THE VARIANCE ARCHITECTURE APPROACH TO THE STUDY OF CONSTRUCTS IN ORGANIZATIONAL CONTEXTS

A dissertation presented to

the faculty of

the College of Arts and Sciences of Ohio University

In partial fulfillment

of the requirements for the degree

Doctor of Philosophy

Dan J. Putka

June 2002
THE VARIANCE ARCHITECTURE APPROACH TO THE STUDY OF CONSTRUCTS IN
ORGANIZATIONAL CONTEXTS

BY

DAN J. PUTKA

This dissertation has been approved
for the Department of Psychology
and the College of Arts and Sciences by

Jeffrey B. Vancouver
Associate Professor of Psychology

Leslie A. Flemming
Dean, College of Arts and Sciences
The purpose of this dissertation is to introduce a novel meta-methodology for studying constructs in organizational contexts based on archival data. The variance architecture approach is a guided exploratory research method designed to estimate (a) the dispersion of variance in constructs across the multiple facets on which they commonly vary and (b) the stability of that dispersion across contexts and time. The approach has its roots in Cattell's (1966) Basic Data Relations Matrix as well as Cronbach, Rajaratnam, and Gleser's (1963) Generalizability theory (G-theory). The General Linear Mixed Model provides the basis for estimating the dispersion of variance in constructs across multiple facets, whereas meta-analytic methods are used to evaluate the stability of that dispersion across contexts and time. An in-depth discussion of the approach is provided, and an example of applying it to the Organizational Citizenship Behavior (OCB) construct is undertaken. By examining OCB through the lens of the variance architecture approach, several questions pertinent to its study are answered, areas where current OCB research is lacking are identified, and both the utility and limitations of this approach are illustrated.

Approved  Jeffrey B. Vancouver

Associate Professor of Psychology
Acknowledgments

My sincerest thanks go to my family, friends and colleagues at Ohio University, DDI, and HumRRO who played an instrumental role in seeing this project to an end. First and foremost, I want to thank my wife Juliane, without whose understanding and patience this dissertation would have never been competed. My parents Joe and Gloria Putka, for serving as shining examples of the success that can be achieved in life through hard work, loving nurturance of one’s children, and faith in God. My girls, Alexa and Sarah, for providing me with the inspiration and smiles to see this to an end. Jeff Vancouver, my academic advisor and mentor, for enduring multiple drafts over the past several years, and an endless stream of quick questions. My dissertation committee (Claudia Gonzalez-Vallejo, Paula Popovich, Rick Milter, Bruce Carlson, and Jeff Vancouver), for taking the time to review this work and offer thoughtful commentary. Bruce Carlson and Charlie Thompson, for ensuring that I had the computer resources to carry out this project. Cindy Poole, for facilitating my interaction with the technical support staff of SAS. My classmates in the Psychology Department at Ohio University, for taking the time to listen and provide support. My friends in the Athens community, in particular Dave and Michelle Miller, and Mike and Julie Finney, for your support and friendship throughout. Last, but not least, my colleagues at DDI and HumRRO, for your encouragement and understanding as I brought this project to an end. Thanks again to all of you for helping my dream become reality.
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>4</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>5</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>6</td>
</tr>
<tr>
<td>List of Tables</td>
<td>9</td>
</tr>
<tr>
<td>List of Figures</td>
<td>10</td>
</tr>
<tr>
<td>Introduction</td>
<td>11</td>
</tr>
<tr>
<td>The Present Study</td>
<td>14</td>
</tr>
<tr>
<td>An Overview of the Variance Architecture Approach</td>
<td>15</td>
</tr>
<tr>
<td>Historical Roots of the Variance Architecture Approach</td>
<td>16</td>
</tr>
<tr>
<td>Cattell's BDRM</td>
<td>17</td>
</tr>
<tr>
<td>Cronbach et al.'s (1963) Generalizability Theory</td>
<td>19</td>
</tr>
<tr>
<td>Comparison of the BDRM, G-theory, and the VAA</td>
<td>21</td>
</tr>
<tr>
<td>Summary</td>
<td>24</td>
</tr>
<tr>
<td>The Content of the Variance Architecture Approach</td>
<td>25</td>
</tr>
<tr>
<td>Constructs Arising in Organizational Contexts</td>
<td>25</td>
</tr>
<tr>
<td>Formulating the Common Facets of the Variance Architectures Approach</td>
<td>26</td>
</tr>
<tr>
<td>Facet Descriptions</td>
<td>30</td>
</tr>
<tr>
<td>Environmental Facets of Variation</td>
<td>30</td>
</tr>
<tr>
<td>Evidence for Environment-Related Variance</td>
<td>31</td>
</tr>
<tr>
<td>Defining Units of the Environmental Facets</td>
<td>32</td>
</tr>
<tr>
<td>Measurement Facets of Variation</td>
<td>34</td>
</tr>
<tr>
<td>Evidence for Measurement-Related Variance</td>
<td>34</td>
</tr>
<tr>
<td>Defining Units of the Measurement Facets</td>
<td>36</td>
</tr>
<tr>
<td>Summary of VAA Facets</td>
<td>38</td>
</tr>
<tr>
<td>Benefits of the Variance Architecture Approach</td>
<td>38</td>
</tr>
<tr>
<td>Identifying the Relative Contribution of Different Sources of Variance in Constructs</td>
<td>39</td>
</tr>
<tr>
<td>Establishing Upper Bounds of Predictability</td>
<td>40</td>
</tr>
<tr>
<td>Clarifying Appropriate Levels of Analysis for Facets with Loosely Defined Units</td>
<td>41</td>
</tr>
<tr>
<td>Identifying the Stability of Facet Contributions</td>
<td>42</td>
</tr>
<tr>
<td>Comparing Different Constructs</td>
<td>43</td>
</tr>
<tr>
<td>Comparing Different Methods of Measurement</td>
<td>46</td>
</tr>
<tr>
<td>Ancillary Benefits of the VAA</td>
<td>47</td>
</tr>
<tr>
<td>Summary of VAA Benefits</td>
<td>48</td>
</tr>
<tr>
<td>The Variance Architecture Approach Research Process</td>
<td>49</td>
</tr>
<tr>
<td>Gathering the Data</td>
<td>50</td>
</tr>
</tbody>
</table>
Laying Out Architectural Designs ................................................................. 53
Assessing the Structure and Adequacy of Data for VAA Analyses .............. 54
Laying Out Architectures in Varying Levels of Detail ..................................... 55
Addressing Confounded Sources of Variation in the VAA ............................... 56
Summary ........................................................................................................ 60
Modeling Variance via the General Linear Mixed Model ................................. 61
An Overview of the GLMM ............................................................................. 61
Variance in the GLMM ....................................................................................... 64
Examples of the Decomposition Process .......................................................... 66
Estimating Model Parameters ........................................................................... 78
Model Fitting Issues in the VAA ................................................................. 86
Meta-Analytically Assessing Architecture Stability ........................................... 93
Comparing Variance Architectures ............................................................... 96
Architecture-Level Comparisons ..................................................................... 98
Facet-Level Comparisons ............................................................................... 103
Summary of Variance Architecture Comparisons ......................................... 105
The Variance Architecture Approach: Summary ............................................. 107
Organizational Citizenship Behavior through a VAA Lens ............................... 109
Definitions and Examples of OCB ................................................................. 109
The Importance of Examining OCB ............................................................... 111
OCB-Related Research Questions ................................................................. 113
The Adequacy of VAA-related Data in the OCB Literature ......................... 114
Architectures of the OCB Construct in General .......................................... 115
Architectures of OCB Dimensions .............................................................. 116
Architectures of OCB Revealed by Different Measurement Methods .......... 119
Architectures of OCB through a Multitrait-Multimethod Lens ...................... 123
Summary ....................................................................................................... 125
Method ........................................................................................................ 126
Identification of Data Sets for Inclusion in the Present Study ......................... 127
Establishing Criteria for Inclusion ............................................................... 128
Procedures for Acquiring Raw Data Sets ..................................................... 129
Overview of Data Collection Results ............................................................ 130
Laying out Architectures for OCB based on the Data Available .................... 141
General Analysis Strategy ............................................................................. 144
Results ....................................................................................................... 148
Descriptive Information and Normality Checks ............................................. 148
Self-Ratings .................................................................................................. 148
Peer Ratings ................................................................................................. 149
Supervisor Ratings ....................................................................................... 151
Basic Variance Overview ............................................................................. 153
Delineating Variance Architectures Underlying Each Dimension of OCB .... 158
Self-Ratings .................................................................................................. 159
Peer Ratings ................................................................................................. 163
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supervisor Ratings</td>
<td>167</td>
</tr>
<tr>
<td>Addressing Architecture-Level and Facet-Level Research Questions</td>
<td>172</td>
</tr>
<tr>
<td>Architecture-Level Comparisons- Relative Facet Contributions</td>
<td>172</td>
</tr>
<tr>
<td>Architecture-Level Comparisons- Stability of Relative Facet Contributions</td>
<td>176</td>
</tr>
<tr>
<td>Facet-Level Comparisons- Absolute Facet Contributions</td>
<td>179</td>
</tr>
<tr>
<td>Facet-Level Comparisons- Stability of Relative Facet Contributions</td>
<td>182</td>
</tr>
<tr>
<td>Discussion</td>
<td>184</td>
</tr>
<tr>
<td>Insights for OCB Research</td>
<td>184</td>
</tr>
<tr>
<td>Data Gathering, Coding, and Measurement Designs</td>
<td>184</td>
</tr>
<tr>
<td>The Current State of the Variance Architectures Underlying OCB</td>
<td>186</td>
</tr>
<tr>
<td>Future Implementation of the Variance Architecture Approach</td>
<td>189</td>
</tr>
<tr>
<td>Availability of VAA-Related Data</td>
<td>189</td>
</tr>
<tr>
<td>Computational Issues in the VAA</td>
<td>191</td>
</tr>
<tr>
<td>Refining Methods for Comparing Variance Architectures</td>
<td>193</td>
</tr>
<tr>
<td>Next Steps</td>
<td>195</td>
</tr>
<tr>
<td>A Final Note</td>
<td>197</td>
</tr>
<tr>
<td>References</td>
<td>198</td>
</tr>
<tr>
<td>Appendix A: Software and Hardware Issues Considerations in the VAA</td>
<td>227</td>
</tr>
<tr>
<td>Appendix B: Formulas for Confidence Intervals around Facet Contributions</td>
<td>230</td>
</tr>
<tr>
<td>Appendix C: Formulas for Estimating the Stability of Facet Contributions</td>
<td>232</td>
</tr>
<tr>
<td>Appendix D: Initial Letter Sent to Researchers for Raw OCB Data</td>
<td>234</td>
</tr>
<tr>
<td>Appendix E: Follow-Up Letter Sent to Researchers for Raw OCB Data</td>
<td>237</td>
</tr>
<tr>
<td>Appendix F: List of OCB Scale Items Examined by Dimension</td>
<td>238</td>
</tr>
<tr>
<td>Appendix G: Formulas for Estimating Random Effects of Each Facet</td>
<td>239</td>
</tr>
<tr>
<td>Endnotes</td>
<td>241</td>
</tr>
</tbody>
</table>
List of Tables

Table 1. Comparison of the BDRM, G-theory and the VAA .................................................... 22
Table 2. The Structural Content of a Construct's Variance Architecture ........................................ 38
Table 3. Benefits and Sample Research Questions Stemming from the VAA ................................. 49
Table 4. A Summary of the Eight Basic Types of Variance Architecture Comparison .................. 106
Table 5. Summary of OCB-Related Research Questions .......................................................... 126
Table 6. Results of the Data Collection Process ........................................................................ 131
Table 7. Data Sets Examined in the Present Study ...................................................................... 134
Table 8. Summary of Variance Architectures Examined in the Present Study .......................... 145
Table 9. Descriptives and Normality Checks for Self-Ratings .................................................. 149
Table 10. Descriptives and Normality Checks for Peer Ratings ............................................... 150
Table 11. Descriptives and Normality Checks for Supervisor Ratings ...................................... 152
Table 12. Basic Variance Overview by Rating Method .............................................................. 154
Table 13. Comparing the Fit of Architectures with and without Dimension Information ............. 156
Table 14. Comparing the Fit of Architectures Based on Self-Ratings ......................................... 159
Table 15. Variance Architecture Summary for Self-Ratings ...................................................... 161
Table 16. Comparing the Fit of Architectures based on Peer Ratings ....................................... 163
Table 17. Variance Architecture Summary for Peer Ratings ...................................................... 164
Table 18. Comparing the Fit of Architectures based on Supervisor Ratings ............................ 167
Table 19. Variance Architecture Summary for Supervisor Ratings .......................................... 169
Table 20. An ASI-RC-based MTMM Matrix .............................................................................. 173
Table 21. An ASI-RS-based MTMM Matrix .............................................................................. 177
Table 22. Computation Times for Architecture Analyses in the Present Study (hrs:min:sec) ....... 192
List of Figures

Figure 1. Map of Cattell’s (1966) BDRM Facets onto the Common Facets of the VAA ............. 28
Figure 2. A Three-dimensional Framework for Comparing Variance Architectures .................. 97
Figure 3. A Summary of the Variance Architecture Approach to the Study of Constructs ....... 108
Figure 4. Absolute Facet Contributions as Revealed by Different Rating Methods ............... 180
Introduction

The questions of person vs. environment, nature vs. nurture, and true-score vs. error, all speak to the fundamental question of sources of variance in constructs. Generally, these classic debates are bipolar (i.e., one vs. another); yet, in applied settings, these issues may be more complex and layered. The organizational context in particular provides a rich multifaceted environment across which individuals’ behaviors, cognitions, and affective states vary. For example, past research indicates that constructs arising in organizational contexts (e.g., job satisfaction, deviance, work values) covary not only with individual difference variables, but also with variables stemming from higher-level factors, such as (a) the units of organization in which individuals function (e.g., organizations, departments, groups, jobs), (b) the cultural / societal context in which the organization is embedded, and (c) the interaction between these factors (Bliese, 2000; Chao, 2000; Hattrup & Jackson, 1996; Kozlowski & Klein, 2000; Roberts, Hulin, & Rousseau, 1978). Such relationships imply that constructs vary not only across individuals, but also across facets of the organizational and cultural environment (i.e., there are mean differences in constructs across organizations, groups, jobs, and cultures).

In addition to facets of the organizational and cultural environment, variance in constructs also arises from conditions surrounding their measurement (Feldt & Brennan, 1989). Just as the organizational context provides a rich multifaceted environment across which individuals’ behaviors, cognitions, and affective states vary, it also provides for the opportunity to use richly faceted measurement designs in assessing individuals’ standing on various constructs. Measurement designs employed the Industrial/Organizational Psychology (I/O) literature have typically consisted of such facets as (a) observers (e.g., different supervisors or peers providing ratings), (b) items (e.g., different items on a personality inventory), (c) stimuli (e.g., different exercises used to evoke an individual’s standing on a construct), and (d) occasions (e.g., different
times when the construct is assessed) (Hoyt & Kerns, 1999; Schmidt & Hunter, 1996). The psychometric literature has long viewed these common facets of measurement designs as potential sources of variance (Hoyt, 2000; Lord & Novick, 1968). Thus, at least part of the observed variance in a construct is likely to arise from the conditions that surround its measurement.

Though the multifaceted nature of construct variance is often recognized in the I/O literature (e.g., Klein & Kozlowski, 2000), rarely are more than two facets of variation considered simultaneously in studies that examine a given construct. Typically, within any single study conducted in the I/O literature, both practical and organizational constraints, as well as theory limit/dictate the facets on which information is available (i.e., by influencing one’s sampling and measurement designs). To the extent that researchers in a particular area of study concern themselves with testing theory focused on single facets of variation only (e.g., individual-level variation), information regarding variation of the construct across other facets may be lacking.

For example, studies of deviant behavior at the individual-level that are conducted within single organizations cannot determine the degree to which deviance varies across organizations. Studies of deviant behavior at the organizational-level that are conducted across organizations cannot determine the degree to which deviance varies across individuals if only organization-level measures of deviance are assessed (e.g., base rates of theft; Wimbush & Dalton, 1997). Studies that examine group-level deviance where individuals are nested within groups (i.e., each individual works in only one group), as opposed to crossed with groups (i.e., each individual works in multiple groups), cannot separate individual-level variance and variance arising from individual-by-group interactions (e.g., Robinson & O'Leary-Kelly, 1998). Although the aforementioned examples of studies often provide insight as to covariates of deviance for the single source of variation on which they focus (e.g., individual-level, organization-level), they do
not provide good indicators of where the totality of variance in deviance primarily lies (e.g., at the individual-level or organization-level). To provide such indicators, multiple facets of potential variation must be examined simultaneously.

A recent trend in many areas of the I/O literature has been on the formation of multilevel theory (e.g., Kozlowski & Klein, 2000). Multilevel theories explicitly recognize that variance in constructs is a function of variables stemming from multiple “levels” (i.e., facets of variation). To test such theories, multilevel researchers often sample across multiple facets of the organizational environment (e.g., Griffin, 1997; Kidwell, Mossholder, & Bennett, 1997). Thus, multilevel studies typically have broader sampling designs than studies that focus on single-level theories. Nevertheless, sampling designs in single multilevel studies remain limited by both practical and organizational constraints, as well as theory.

Theory may limit the samples obtained in multilevel studies in that sampling of units usually occurs only across facets that are pertinent to one’s multilevel theory (e.g. examining variation in a construct as a function of only individual-level and group-level variables; George, 1990). Practical constraints often lead to problems in obtaining large samples of units from higher-level facets in the design (e.g., groups, organizations; Raudenbush & Liu, 2000). Organizational constraints often lead to the examination of nested as opposed to crossed facets (which leads to difficulty in separating variance arising from single facets and interactions; Shavelson & Webb, 1991). For example, in studies of job performance, one often has direct supervisors rate their subordinates. Typically, individuals have only one direct supervisor. Based on this measurement design (i.e., individuals nested within observers), one will not be able to uniquely estimate variance attributable to individuals, or to the individual-by-observer interaction (Hoyt, 2000). Thus, within both single-level and multilevel research studies, researchers are
limited in their ability to make generalizable statements about the degree to which a construct varies across multiple facets based solely on the results of their single study.

The Present Study

A general conclusion that can be drawn from the discussion above is that variation in constructs occurs across more than simply individuals. Indeed past research has demonstrated that many constructs of interest to organizational researchers vary across (a) individuals, (b) different facets of the organizational context (e.g., jobs, groups, organizations), (c) different cultures, (d) different facets of the measurement design used to assess individuals’ standings on a construct, and (e) interactions between each of these facets of variation. In later sections, I cite specific examples of such research, and build upon the insight they provide to formulate a set of common facets that can serve as a foundation for the exploration of variance in any construct that arises in organizational contexts. Although research and theory suggest variation exists across many facets, they do not indicate the relative contribution of each of these sources of variation to the total variance in any given construct. Nor does any systematic methodology currently exist in the literature for doing so.

The key to providing estimates of the relative contribution of each facet of variation is to consider data across multiple studies simultaneously. Only through conducting a systematic investigation of variance in constructs that is meta-analytic in scope will one be able to overcome the theoretical, practical and organizational constraints often characteristic of the sampling and measurement designs used in single studies. By pooling data across studies, one will be able to investigate several facets of variation simultaneously, and generate stable estimates for the relative contribution of each facet to total variance in a construct. In this way, the research method introduced in this dissertation is epistemologically similar to traditional meta-analysis (e.g., Hunter & Schmidt, 1990), in that knowledge about a particular construct is gained through
empirical aggregation of past research. As later sections reveal, the research method introduced here shares other similarities with meta-analysis as well (e.g., data gathering procedures, decision processes, sampling concerns, statistical techniques; Wanous, Sullivan, & Malinak, 1989). Nevertheless, the methodology introduced in this paper is quite distinct from traditional meta-analysis, and those differences will become quite apparent as the method is introduced.

In light of the observations made above, my purpose in this dissertation is fourfold. I will (a) describe a guided exploratory research methodology for systematically delineating the contribution of several common facets of variation² to the total variance in any given construct arising in organizational contexts through the aggregation of archival data gathered in the applied literature, (b) provide a framework for examining and contrasting patterns of variance underlying such constructs, (c) detail how obtaining information regarding the dispersion of variance underlying a construct may benefit both organizational science and practice, and (d) provide an empirical example of applying this method to archival data obtained from the applied literature. I will refer to this methodology as the variance architecture approach (VAA) to the study of constructs in organizational contexts.

An Overview of the Variance Architecture Approach

The variance architecture approach (VAA) can simply be defined as a guided exploratory research methodology for delineating the variance architectures of constructs. The variance architecture (VA) of a construct is defined as the dispersion of variance in a construct across the common facets on which it may vary, and the stability of that dispersion across contexts and time. The set of facets that constitute a construct's variance architecture comprise sources of individual, environmental, and measurement-related sources of variation. These facets serve as the foundation of the VAA and can give researchers a common point of departure for using the VAA
as a guided exploratory method for delineating where the totality of variance in a construct resides (Cattell, 1966).

Delineating the variance architectures of constructs can be quite beneficial to organizational science and practice in that it can: (a) identify areas where future theory-building, prediction, and intervention efforts might have most / least impact for a given construct, (b) aide in estimating likely upper bounds of variance accounted for by predictors stemming from a certain level for a given construct, (c) clarify appropriate levels of analysis for a given construct, (d) identify the degree of stability in the dispersion of variance across facets in a given construct, (e) provide a unified framework for comparing different constructs, methods of measuring a single construct, and dimensions of a multidimensional construct across multiple levels of analysis, and (f) identify areas of the research literature that are scant with regard to the contribution a facet makes to construct variation. These benefits will be fully described in later sections.

In the sections that follow, I outline the variance architecture approach in detail. Specifically, the following sections detail (a) the historical roots of the approach in the scientific literature, (b) the content of the approach, in terms of the common facets that constitute constructs’ variance architectures and the focal constructs of interest in the VAA, (c) the benefits of adopting the VAA for organizational science and practice, and (d) the VAA research process itself. After outlining the VAA in detail, I provide an illustrative empirical example of the approach using archival data gathered on the organizational citizenship behavior construct (Organ, 1988).

**Historical Roots of the Variance Architecture Approach**

Frameworks for conceptualizing the multifaceted nature of variance underlying constructs have primarily emerged from two areas, experimental psychology and psychometrics.
Two general frameworks that are particularly relevant to the present exposition of the variance architecture approach are Cattell's (1966) Basic Data Relations Matrix (BDRM) and Cronbach et al.'s (1963) Generalizability theory (G-theory). The perspectives on construct variance that these frameworks offer differ greatly in terms of their original purpose, specificity, and focus, yet both provide a historical grounding for subsequent discussion of the VAA. The following sections describe the perspectives these frameworks offer and how they compare to the perspective offered by the VAA.

Cattell's BDRM

Cattell (1966) outlined a framework for examining all possible relationships between constructs as a function of ten facets. The scope of the framework presented by Cattell was designed to be broad enough to encompass the entire variety of relationships between constructs that could be examined by psychological researchers. The ten facets he identified consisted of five general facets: persons, stimuli, environmental background, responses, and observers; as well as five facets that represented the crossing of each general facet with time (i.e., state of the person, variant of the stimuli, phase of the environmental background, style of response, and condition of the observer). The "levels" or "coordinates" within these facets were described as falling on nominal scales (e.g., person 1, person 2, person 3), whereas the "attributes" of such coordinates were interval or ratio scaled (e.g., height, weight, and other continuously-scaled characteristics of each person). Cattell (1966) named this systematic delineation of the ways in which two or more constructs may possibly covary the Basic Data Relations Matrix (BDRM), or more simply, the “data box.” He described the BDRM as "the scientist's complete box of data - of all facts and relations available for study" (p. 80).

Cattell (1966) argued that the BDRM provided scientists (a) an aide in forming more encompassing statistical models of the theories used to tie constructs together, (b) a manner of
checking that the theory proposed is internally consistent, and (c) a means for researchers to
discover new ways in which the relationships between constructs may be stretched that would
otherwise go unrecognized without the existence of such a comprehensive framework. Alluding
to this last claim, Coan (1961) stated that by considering the totality of the number of facets on
which relationships between constructs may vary, the BDRM's "major virtue…is that it can
suggest forms of research which might otherwise be overlooked" (from Cattell, 1966, p.71).

Because the purpose of Cattell's (1966) exposition of the BDRM was to delineate the
totality of possible relations that could exist between multiple constructs, decomposing the
variance in single constructs was not the focus of his work (unlike the VAA). Nevertheless,
Cattell (1966) did describe how the BDRM provided a framework for decomposing the variance
in a single construct as a function of the 10 general facets he introduced in his work. The
methodology he espoused for decomposing the variance was based on analysis of variance
(ANOVA) methodology. Specifically, Cattell (1966) suggested that one could estimate the
amount of variance in a given attribute (i.e., construct) that is associated with each of his facets as
well as their interactions, via their sums of squares arising from a traditional ANOVA model.
Although, such ANOVA-based methods for decomposing variance have merit, Cattell (1966) did
not pursue the problems that employing such methods to decompose variance might have in any
substantial detail. As will be discussed later, some characteristics of ANOVA-based
decomposition methods have limitations that make them difficult to apply as general variance
decomposition tools for use in the VAA.

Even though the methods proposed by Cattell (1966) for delineating the contribution of
each of the facets in his general framework to the total variance in a single construct are limited,
the benefits he cited for achieving such a decomposition of variance remain true. Specifically,
Cattell (1966) stated that the value of using the BDRM as a framework for decomposing variance in a given construct was:

"(1) its reminding us of the totality of sources of variance and interaction; (2) its offering of a basis of real variance sources into which operationally to analyze proposed conceptual sources of variance; (3) its suggestion of fresh conceptual resources; (4) its definition of error in a number of alternative ways; (5) its offering of a perspective on the relative importance and meaning of a variety of variance sources" (p. 116).

As later sections will detail, given the primary focus of the VAA is on large-scale decomposition of the variance in constructs along a set of general set of common facets, the benefits Cattell (1966) cited above apply equally as well to the VAA.

_Cronbach et al.’s (1963) Generalizability Theory_

In his discussion of the decomposition of variance in constructs, Cattell (1966) alluded to the relative nature of the labels "error" variance and "true-score" variance in construct measurement and in doing so paralleled arguments that were concurrently being made in the Generalizability theory (G-theory) literature (Cronbach et al., 1963). Specifically, Cattell (1966) made the observation, that which facets or interaction of facets in his framework were considered error variance in a given study was dependent on the types of relationships the researcher was studying and the conclusions that researcher wished to draw. This observation is a key premise underlying G-theory (Brennan, 1983; Shavelson & Webb, 1991).

Unlike the perspective on construct variation provided by the BDRM, G-theory does not offer a fixed set of facets along which relationships between multiple constructs can be classified, nor even a fixed set of facets along which variance in any single construct can be decomposed. Rather, the focus of G-theory is on estimating the dependability with which a person’s standing
on a given construct is assessed as a function of the facets in one’s measurement design (e.g., items, observers, occasions; Cronbach et al., 1963). One of the main characteristics that differentiates G-theory from classical test theory notions of reliability is that G-theory recognizes and accounts for multiple sources of variation in measurements simultaneously (i.e., besides simply true-score variance and one source of error variance; Feldt & Brennan, 1989).

Like Cattell (1966), Cronbach and his colleagues (1963) recognized the multifaceted nature of variance underlying the measurement of constructs. However, Cronbach and his colleagues (1963) did not stop there, they developed an entire psychometric theory around their multifaceted perspective of measurement-related variance. Developing G-theory necessitated that Cronbach and his colleagues (1963) take a much more active approach to how the variance in constructs could be decomposed. This active approach was necessary because the components of variance resulting from such decompositions served as the basis for the centerpieces of their theory, namely generalizability coefficients.

The method of variance decomposition espoused in G-theory is based on a random-effects ANOVA model (Jackson & Brashers, 1994). Unlike Cattell’s BDRM and the VAA, variance decomposition within G-theory is entirely dependent on the facets that constitute the measurement design used to assess the given construct (e.g., items, occasions, observers). Within the context of the random-effects ANOVA model, each facet of a measurement design is treated as a random factor. The variance in the measure of the construct can be decomposed into several variance components (i.e., an estimate of variance accounted for by a given facet), with one variance component corresponding to each of facet in the design (e.g., person-facet, item-facet) and each of their interactions (e.g., person-by-item interaction).

Although the random-effects ANOVA model can result in a very parsimonious decomposition of construct variance, it has some serious limitations that restrict its usefulness as
a general variance decomposition technique. For example, it requires that one work with a balanced data set (e.g., no unbalanced nesting or missing data) to achieve unbiased estimates of one's variance. Unfortunately, researchers employing G-theory rarely work with balanced data (e.g., Greguras & Robie, 1998). In such situations researchers must rely on alternative estimators of variance components besides the expected means squares estimators (EMS) arising from traditional random-effects ANOVA analyses, lest they have to randomly delete data to balance their design (Brennan, 1983).

Although, the issue of unbalanced data is typically dealt with using alternative estimators of variance components that are unbiased in the face of unbalanced data (e.g., REML, ML, mivque0; Searle, Casella, & McCulloch, 1992), another more serious limitation of the statistical model underlying G-theory is that it is constrained by assumptions necessary to make it applicable for its intended purpose (i.e., for calculating coefficients of generalizability). Although making such assumptions are necessary for the tenets of G-theory to hold and to generate unbiased generalizability coefficients (Bost, 1995; Smith & Luecht, 1992), as will be discussed in later sections, such assumptions may not reflect the true nature of variation within the construct. Because the VAA’s main purpose is not to generate G-coefficients, but rather to reflect the true nature of variation within the construct, a more general model of variance decomposition will be used within the VAA, namely the general linear mixed model (GLMM; Littell, Miliken, Stroup, & Wolfinger, 1996).

Comparison of the BDRM, G-theory, and the VAA

The perspectives on variance offered by the BDRM, G-theory, and VAA share some notable similarities and differences. Table 1 illustrates a side-by-side comparison of various features of the three frameworks along several different dimensions.
Table 1. *Comparison of the BDRM, G-theory and the VAA*

<table>
<thead>
<tr>
<th>Dimension</th>
<th>BDRM</th>
<th>G-theory</th>
<th>VAA</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Purpose</td>
<td>Forming a taxonomy of potential relations between all constructs</td>
<td>Reliability estimation</td>
<td>Delineating variance architectures of constructs arising in organizational contexts</td>
</tr>
<tr>
<td>Focal Constructs</td>
<td>All constructs</td>
<td>All constructs</td>
<td>All constructs arising in organizational contexts</td>
</tr>
<tr>
<td>Study Type</td>
<td>Conceptual (not empirical)</td>
<td>Single empirical study</td>
<td>Meta-analytic</td>
</tr>
<tr>
<td>Data Source</td>
<td>No data (not empirical)</td>
<td>Data stems from the measurement of a construct in a single study</td>
<td>Data stems from the measurement of a construct in several past studies (archival)</td>
</tr>
<tr>
<td>Statistical Model</td>
<td>ANOVA (conceptual)</td>
<td>Random-effects ANOVA</td>
<td>GLMM</td>
</tr>
<tr>
<td>Facet Content</td>
<td>Fixed</td>
<td>Variable</td>
<td>Fixed</td>
</tr>
<tr>
<td>Facet Scope</td>
<td>General enough to be applicable to all constructs</td>
<td>Entirely dependent on the measurement design in a given study</td>
<td>Applicable to constructs arising in organizational contexts</td>
</tr>
</tbody>
</table>

The first dimension along which the BDRM, G-theory and the VAA are compared in Table 1 concerns their general purpose. As stated earlier, the purpose of the VAA is to delineate the (a) relative dispersion of variance in a construct across the common facets on which it may vary and (b) stability of that dispersion across contexts and time (i.e., delineate its variance architecture). In this regard, the VAA is more similar to G-theory than the BDRM in that both VAA and G-theory can be viewed as active approaches to variance decomposition (i.e., both frameworks provide explicit methods for decomposing variance in a given construct), where as the BDRM simply provides a set of facets along which variance may be decomposed.

A second dimension along which the three variance frameworks can be compared concerns their constructs of focal interest. Whereas the focus of both the BDRM and G-theory is
on all constructs in general, the VAA is primarily concerned with constructs that *arise in the organizational contexts*. Constructs that *arise in organizational contexts* are constructs that either (a) arise as a result of persons’ functioning in the organizational context (e.g., job performance, job satisfaction, work stress, job knowledge), or (b) have implications for organizational outcomes but do not necessarily arise as result of persons’ functioning in the organizational context (e.g., intelligence, conscientiousness, negative affectivity, need for achievement).

The third and fourth dimensions along which these frameworks may be compared concerns the type of study in which these frameworks are likely to arise, and the source of data for such studies. Because the BDRM was primarily introduced as a conceptual framework, empirical decomposition of variance was not its intended purpose. Thus, the “study type” and “data source” dimensions do not apply well to the BDRM. On the other hand, both G-theory and the VAA proscribe clear methods for decomposing variance to achieve their given purposes. In G-theory, measurement data is typically examined on a study-by-study basis, and a unique G-coefficient is estimated for each study (Shavelson & Webb, 1991). Thus, variance decomposition in the G-theory literature primarily occurs based on data from a single study. Under the VAA, variance decomposition is based on data from several past studies of the focal construct (i.e., data in the VAA is primarily archival). Thus, relative to studies stemming from the G-theory literature, a VAA-type study can be viewed as more of a meta-analytic investigation of the variance underlying a construct.

The fifth dimension along which these frameworks can be compared concerns the statistical model underlying each framework. As stated above, G-theory and the VAA both offer an explicit set of methods for decomposing variance within a construct, as well as a specific statistical model on which that decomposition is based. Cattell's (1966) exposition of the BDRM on the other hand provides little detailed guidance in either regard. Although G-theory and the
VAA share this commonality, the purposes behind decomposing variance in each framework differ. These differences in purpose lead to differences in the statistical models used in each framework. Because the purpose of G-theory is the estimation of G-coefficients, the statistical model it employs to decompose variance is constrained by the assumptions underlying G-theory. Because the focus of the VAA is on delineating variance architectures rather than estimating G-coefficients, the statistical model underlying the VAA need not be constrained by G-theory assumptions. Thus, a much more general statistical model underlies variance decomposition in the VAA (i.e., the GLMM).

The sixth and seventh dimensions along which these frameworks can be compared concerns the nature of facets of interest in each framework. The VAA and BDRM can both be distinguished from G-theory in that both provide a fixed set of facets along which the variance in a construct can be described. G-theory does not limit the facets that it can examine. Although the VAA and BDRM share this commonality, they differ in regard to their facets of interest. The facets of interest in the BDRM are the 10 facets outlined by Cattell (1966) (see above). The facets of interest in the VAA arise from the (a) organizational and cultural context in which an individual functions, (b) the conditions of measurement, (c) interactions among these facets. Thus, the facets examined in the VAA are much more specific to constructs arising in organizational contexts, rather than all psychological constructs in general. The specific facets that are of interest in the VAA will be formally introduced in later sections.

Summary

Although both Cattell's (1966) BDRM, and Cronbach et al.’s (1963) G-theory provide useful perspectives on construct variance, neither fulfills the intended purpose of the VAA approach (i.e., the delineation of constructs’ variance architectures). In no way is this a criticism of either Cattell's BDRM or G-theory, as both were conceived for a particular purpose, and both
fulfill those purposes very well. Nevertheless, the perspectives on construct variance offered by these frameworks provide a solid basis for further discussion of the VAA.

The Content of the Variance Architecture Approach

Content in the context of the VAA carries two meanings. First, there is content in terms of the constructs that are of focal interest in the VAA, namely constructs that arise in organizational contexts (defined earlier). Second, there is content in terms of the common facets of variation that provide the basis for fleshing out the variance architecture underlying a particular construct.

*Constructs Arising in Organizational Contexts*

In addition to focusing only on constructs that arise in organizational contexts, the constructs that this initial exposition of the VAA will focus on is further limited in two ways. First, my initial focus will be on individual-level constructs. An *individual-level construct* is simply defined as a property or characteristic of a person (e.g., attitudes, skills, abilities, cognitions, values, beliefs, behaviors, and affect). Although the VAA may be equally applicable for examining higher-level constructs (e.g., organizational climate, group-cohesiveness), the body of research surrounding these constructs is not nearly as well developed as the literature surrounding the many individual-level constructs of interest in I/O psychology.

Second, the focus of this initial exposition of the VAA will be on constructs that have primarily been operationalized as continuously-scaled variables in the applied literature. The reason for the focus on continuously-scaled variables stems from the fact that statistical models underlying decomposition of variance are developed to a greater degree for continuously-scaled than for categorically-scaled variables (Searle et al., 1992; Wolfinger, 1992). As Searle and colleagues (1992) summarize: "Techniques for the estimation of variance components from binary (0/1) or discrete (categorical) data are much less widely developed than for continuous
data. The lack of methods for such data is due in large part both to the difficulty of specifying realistic models, and once specified, to their computational intractability" (p.367).

Formulating the Common Facets of the Variance Architectures Approach

Prior to introducing the facets of variation that provide a basis for the VAA, it will be useful to introduce a brief analogy. This analogy draws parallels between the architecture of physical structures and the variance architectures underlying constructs arising in organizational contexts. Just as variability in all physical structures can be described along a common set of dimensions (e.g., length, width, height, weight); the variance in nearly all constructs arising in organizational contexts can be described along a set of common dimensions (i.e., their common facets of variation). Furthermore, just as the common dimensions of physical structures serve as the primary means for describing such structures (and provide multiple points on which those structures can be compared); the common facets of the VAA create a common link among all constructs studied in organizational contexts, and thus allow for their description and comparison across many different points using a common language (i.e., the relative contribution of facets constituting their variance architectures).

Before describing the process by which I arrived at the set of facets that serves as the foundation of the VAA, it will be helpful to clarify a few terms: (a) facets, (b) units of a facet, and (c) unit properties. First, a facet refers to any general source of random variation in a given construct. For example, within a typical multilevel study in the I/O literature, individuals and groups are often the focal facets of interest. These facets represent non-specific sources of variation in the construct in that they do not indicate what property of individuals explain variation across individuals, nor what property of groups explain variation across groups. Continuing with the above example, units of the above facets are the individuals and groups that constitute these facets, respectively (e.g., individuals 1, 2 and 3; groups 1, 2, and 3). Thus, we
describe variation as existing across units of a given facet (e.g., variance in a construct across individuals, or across groups). Lastly, there are specific properties of individuals or groups that may explain variation across individuals, or across groups. These unit-properties often take the form of other constructs or variables (e.g., gender, job stress, group-cohesion) that a researcher believes accounts for the variance across units on the construct of interest (e.g., job satisfaction).

The sources of construct variation represented in the set of facets at the foundation of the VAA are based on abstracting Cattell’s (1966) criteria for delineating the 10 facets he identified for his BDRM, in light of the applied research literature relevant to the study of constructs in organizational contexts (e.g., I/O psychology, organizational behavior, social psychology, personality, cross-cultural psychology, psychometrics). In formulating this set of facets the goal was to identify a set of organizationally relevant facets that were comprehensive (not necessarily exhaustive) in that they were arguably "more fundamental than any other that might be considered, and that they suffice to define a behavioral event" (Cattell, 1966, p.79) or observation in the organizational context. Plainly speaking, if one were to parse up the totality of variance in a construct across all of the facets that constitute the foundation of the VAA and their interactions, there would arguably be little if any variance to explain by adding additional facets or interactions among them. To the extent this goal is achieved, the facets of variation that constitute the foundation of the VAA provide researchers a common point of departure and comprehensive framework for guiding researchers in their exploration of the variation in any organizationally-relevant construct. In light of the aforementioned goal, it is not surprising that the facets that provide the foundation of the VAA, share a number of characteristics. First, they all can be tied directly back to Cattell (1966)’s BDRM facets, either directly or indirectly. Figure 1 maps the facets of Cattell’s (1966) BDRM onto the facets of that provide the basis of the VAA. Second, the sources of variation that these facets tap have generally been shown to be relevant to the study of
a number of constructs of interest to organizational researchers. As Cattell (1966) states in justifying his inclusion for 10 facets (as opposed to a simpler framework): "if these are the true dimensions of independently operating sources of variance in an observed psychological event, the needs of science dictate that we recognize and accept them" (p.78). In sections that follow, I cite empirical evidence for the existence of variation across each of the VAA facets for a number of organizationally-relevant constructs.

<table>
<thead>
<tr>
<th>BDRM Facets</th>
<th>vs.</th>
<th>VAA Facets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Persons</td>
<td>Individual</td>
<td></td>
</tr>
<tr>
<td><strong>Environmental Facets</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Environmental Background</td>
<td>Culture</td>
<td>Organization</td>
</tr>
<tr>
<td></td>
<td>Group</td>
<td>Job</td>
</tr>
<tr>
<td><strong>Measurement Facets</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Responses</td>
<td>Item</td>
<td></td>
</tr>
<tr>
<td>Observers</td>
<td>Observer</td>
<td></td>
</tr>
<tr>
<td>Stimuli</td>
<td>Stimulus</td>
<td>Occasion</td>
</tr>
</tbody>
</table>

**Interactions among Facets**
- Each 2-way interaction among facets
- Each n-way interaction among facets

Figure 1. Map of Cattell’s (1966) BDRM Facets onto the Common Facets of the VAA
In light of the process described above, some researchers may feel the list of VAA facets I have arrived at is incomplete. As I state above, although not necessarily exhaustive, the set of VAA facets is arguably comprehensive. In many instances, facets that are “missing” from this framework (e.g., industry, gender, race), can be subsumed by facets already in the VAA framework. For example, one can argue that many such alleged facets are arguably better thought of as unit-properties (i.e. as opposed to facets) that explain variation across units of the given facet from which they stem. Specifically, industry can very well be viewed as a property of organizations that may explain variation across organizations that constitute the organizational facet. Likewise, gender or race can be viewed as properties of individuals that may explain variation across individuals that constitute the individual facet.

As the facet descriptions presented below reveal, another common characteristic of many of the facets that constitute the VAA have units that can be defined at varying levels of specificity. For example, although the job facet reflects one facet of the VAA, one might choose to define its units as jobs, occupations, or job family. This characteristic allows the set of VAA facets to be much more comprehensive than they appear based solely on the names they were given in Figure 1. Further, having such facets where units are loosely defined offers flexibility to researchers who employ the VAA as a structured, comprehensive framework for exploring the variation in a construct. Such flexibility is valuable because operationalizing units in one particular way, may be more salient than another depending on the construct. For example, a particular construct may have variance across job families, but not across jobs. On the other hand, another construct may have variance across jobs, but not job families. In either case, part of the construct’s total variance stems from the job facet, but in one instance it would be of more interest to define units in of the job facet in terms of job families (the former), and in the other it would be of more interest to define units in terms of jobs (the latter).
Facet Descriptions

As stated above, Figure 1 illustrates how the common facets of the VAA can be mapped onto Cattell’s (1966) BDRM facets. Figure 1 also illustrates how the common facets of the VAA can be broken down into four sets: (a) the individual facet, (b) the environmental facets (i.e., culture, organization, group, and job-facets), (c) the measurement facets (i.e., observer, item, stimuli, occasion-facets), and (d) the interactions among these facets (i.e., each two-way and higher-order interaction among VAA facets). Because the individual facet simply reflects variation across individuals, neither a description of its units (i.e., different individuals), nor its inclusion in VAA, need further justification (e.g., individual-level constructs will necessarily consist of individual-level variation). Thus, in the sections that follow, only the environmental facets and measurement facets will be more fully detailed.

Environmental Facets of Variation

As noted, the organizational context provides a rich, multifaceted environment across which constructs may vary. To capture the richness of that multifaceted environment, it is necessary to make the facets of the BDRM less general and more specific to the organizational contexts. Cattell’s (1966) environmental background facet is broken down into four facets within the VAA: a culture-facet, an organization-facet, a group-facet, and a job-facet. The separation of Cattell’s (1966) environmental background facet into these four particular facets was based on (a) past research that has indicated the existence of job-, group-, organization-, and culture-related variance in a wide variety of construct arising in organizational constructs (reviewed below), (b) theory that suggests that variation in constructs arises from such facets of the environment (e.g., Schneider’s (1987) ASA framework; Salancik & Pfeffer’s (1978) social information processing theory; and Bandura’s (1977) social learning theory) and (c) considerations of the structuring of organizations in general (Dansereau, Alutto., & Yammarino, 1984; Klein & Kozlowski, 2000;
Roberts et al., 1978; Rousseau, 1985). By becoming much more specific in terms of the environmental background facet, the VAA allows for the investigation of facets of variation that are of particular interest to organizational researchers and practitioners.

Evidence for Environment-Related Variance

Much research conducted in I/O psychology and related fields has demonstrated the existence of variance across each of the VAA environmental facets for a wide variety of constructs. For example, several researchers have found that job attitudes (e.g., commitment, satisfaction; Angle, & Perry, 1981; Ostroff, 1992) and work behavior (e.g., deviance, Robinson & O’Leary-Kelly, 1998; citizenship, Podsakosff, Ahearne, & MacKenzie, 1997) cluster by group and organization. Such findings are consistent with theories that suggest the strong role of social influence on individuals’ attitudes and behavior (e.g., social information processing theory, Salancik & Pfeffer, 1978; social learning theory, Bandura, 1977). Other researchers have found that personality variables cluster by group and organization as well (e.g., Barrick, Stewart, Neubert, & Mount, 1998; Schneider, Smith, Taylor, & Fleenor, 1998). These findings are consistent with Schneider's (1987) ASA theory which postulates that individuals will be attracted to, selected by, and remain employed in organizations that have aggregate characteristics similar to their own.

Several areas of the I/O literature also suggest that job-level variables covary with individuals’ standings on various constructs (e.g., various dimensions of job performance, Peters, Fisher, & O'Conner, 1982; Murphy & Shiarella, 1997). For example, if an individual works on a job where he or she has little contact with co-workers, then that individuals' ability to help co-workers with work-related problems is likely to be lower than an individual who works in an environment with much co-worker contact. Thus, base rates of helping behavior in the latter type of jobs may be higher than the former type of jobs. Another example of job-level variation in
constructs stems from the literatures on vocational interests and work values. Studies from these literatures have consistently revealed differences in interest and value profiles of individuals who hold different types of jobs, as well as similarity in the interest and value profiles of those who hold similar jobs (e.g., Gay, Weiss, Hendel, Dawis, & Lofquist, 1971; Lau, & Abarahams, 1972; Munson & Posner, 1980; Strong, 1955). Such findings are consistent with Schneider’s (1987) ASA theory operating at the job-level.

Lastly both the I/O and cross-cultural psychology literature suggest that the cultural context in which organizations are embedded is related to individuals' standing on various constructs (e.g., Earley, 1989; Hofstede, 1980; Moorman & Blakely, 1995; Lam, Hui, & Law, 1999; Lind & Earley, 1992; Schwartz, 1992, 1994). For example, Hofstede (1980) found that work-related values tended to cluster by country (e.g., individualism-collectivism, power distance). Such culture-level differences in work values have been used to help explain differences in employees’ perceptions of the breadth of their job requirements across cultures (e.g., Lam et al., 1999).

**Defining Units of the Environmental Facets**

Although the units of the organization-facet may be relatively straightforward (i.e., each organization is a different unit of the organization-facet), the units of the other environmental facets that constitute the VAA may need further description. For example, culture is not viewed as an individual-difference variable within the VAA, but rather as a higher-level facet that gives rise to several specific cultural-level variables that may influence an individuals’ standing on a variety of organizational constructs (Bond & Smith, 1996; Chao, 2000; Hofstede, 1980; Schwartz, 1994). The units of the culture-facet are loosely defined within the VAA and may reflect the culture of the country in which the unit of the organization under study resides (e.g., a particular plant), or the dominant local / regional culture from which it draws the majority of its
local employee base (Chao, 2000). Given that organizations may reside in one city, one country, one global region, or be multiregional or multinational in nature, different units of the same organization may reside in different regional or national cultures. Culture-facets tapping into country-level and regional-level variation may be included in any given construct’s variance architecture should data be available to provide stable estimates of their contributions.

Like the culture-facet, the units of the group-facet are also loosely defined within the VAA. Specifically, units of the group facet can be different workgroups, teams, departments, business units, or any other group-level units that fall between the individual- and organizational-levels. Although group-level research in I/O psychology has primarily focused on the workgroup- or team-units (e.g., George & Bettenhausen, 1990; Kidwell et al, 1997), some variation in constructs may also be explained by variables stemming from “other group”-units of the organizational hierarchy (e.g., departments; plants) and are in need of exploration (Greenberg, 1990; Rousseau, 1985). Facets tapping into workgroup-, department-, and business unit-related variation may be included in any given construct’s variance architecture should data be available to provide stable estimates of their contributions.

As was the case with the culture- and group-facets, units of the job-facet are also loosely defined within the VAA. The units of the job facet may can defined as different jobs, occupations, job families, or other clusters of jobs such as those identified by common job classification frameworks (e.g., O*Net; Peterson, Mumford, Borman, Jeanneret, & Fleishman, 1999). As was the case with the group-facet, facets tapping into job-, occupation- job family-, and other-job cluster-related variation may be included in any given construct’s variance architecture should data be available to provide stable estimates of their contributions.
Measurement Facets of Variation

Although Cattell’s (1966) environmental background facet is broken down into several different facets in the VAA, the facets that pertain to the measurement of constructs in the VAA retain a fairly one-to-one relationship with the measurement-related facets of the BDRM (see Figure 1). Within the VAA, these facets take the form of (a) an item/behavior-facet (analogous to the BDRM’s response facet), (b) an observer-facet, (c) a stimulus-facet, and (d) an occasion facet.

Evidence for Measurement-Related Variance

Evidence for variation related to the aforementioned facets stems from decades of work in the psychometric and I/O psychology literature (Brennan, 1983; Schmidt & Hunter, 1996). For example, on any test of general cognitive ability it is almost a given that items will vary in difficulty. If items vary in difficulty (or behaviors on a scale have different base rates), then individuals’ mean scores on them will necessarily vary, thus leading to variation across items (Allen & Yen, 1978). Furthermore, because coefficients of inter-item consistency of construct measures rarely equal 1, individual-by-item variation is almost guaranteed to account for a portion of variance in the measurement of a construct (e.g., coefficient alpha decreases as individual-by-item variance increases, holding the number of items constant; Feldt & Brennan, 1989).

The existence of observer-related variation is clear in several areas of the I/O literature (e.g., job interviews, job analysis surveys, job-performance ratings). For example, a large concern among researchers investigating the quality of job performance ratings has been distinguishing between observers’ rating idiosyncrasies and true ratee performance differences (Scullen, Mount, & Goff, 2000). Indeed, researchers commonly find that observers differ in their ratings of the same individual’s performance, even if the observers are from the same level of the
organizational hierarchy (e.g., between-supervisor or between-peer disagreement; Greguras & Robie, 1998). Such disagreement between observers suggests that some of the variance in job performance ratings arise from observer differences.

The existence of stimulus-related variation is seen clearly in the I/O research regarding executive assessment centers (e.g., Bycio, Alvares, & Hahn, 1987; Harris, Becker, & Smith, 1993). For example, Schneider and Schmitt (1992) found that individuals’ relative standing on a given dimension (i.e., an ability-related construct, e.g., oral communication skill), varied as a function of the exercise (e.g., in-basket, role-play) used to evoke it. This finding suggests that variance stemming from an interaction between individuals and stimuli may account for at least some of the variance in the construct being assessed.

Another example of stimulus-related variation stems from the personality literature. For example, some researchers have found support for the theory that a person’s standing on a personality construct is best understood by observing their pattern of behavioral responses in different situations where opportunities to display behaviors consistent with the personality construct may arise (Mischel and Shoda, 1995; Shoda, Mischel, & Wright, 1994). These researchers claim that assessing a person's standing on a given personality construct using a single stimulus may represent only a partial reflection of the person’s true standing on the construct, because it is the pattern of behavioral responses, rather than an isolated response to a single stimuli, through which personality is revealed. Differentiating individual’s personality based on patterns of responses in the face of different stimuli suggests that an individual-by-stimulus interaction accounts for at least some of the variance in personality constructs.

Evidence for occasion-related variation in constructs has typically been provided in two forms. One form of occasion-related variation is in terms of short-term, temporal variation in a persons’ standing on a construct arising as a function of what’s been called transient error (i.e.,
differences in mood, mental efficiency or mental state across occasions; Schmidt, Viswesvaran, & Ones, 2000). Often this form of occasion-related variation is used as the error term in calculations of test-retest reliability (Feldt & Brennan, 1989). The other form of occasion-related variation is other is in terms of changes more systematic changes in an individual’s standing on a construct across time arising from training or development (e.g., skill-acquisition, Ackerman, 1988; job performance, Farrell & McDaniel, 2001).

**Defining Units of the Measurement Facets**

Unlike many of the environmental facets that constitute the VAA, the units of the measurement facets in the VAA can be defined tightly (although occasion is an exception- to this rule- see below). For example, the units of the item-facet consist of the different items or behaviors used to assess an individual’s standing on a construct. The item-facet taps variation across such items and behaviors. The units of the observer-facet consist of different observers or raters of an individual's standing on a construct. The observer-facet taps variation across such observers and raters. The units of the stimulus-facet consist of different exercises or events that may evoke an individual’s standing on a given construct of interest. For example, in the context of assessing oral communication skill, one may administer an individual a number of different exercises that give the individual the opportunity to display their oral communication skill in various contexts (e.g., group meeting, formal presentation, phone conversation). The stimuli in this case would be the different exercises used to evoke the individual’s oral communication skill and variation due to the stimuli-facet would reflect variation in the individuals’ oral communication across the exercises.

Much like the units of the group-, culture-, and job-facets described earlier, the units of the occasion-facet are loosely defined and consist of the different times (hours, days, months, years) at which an individual is assessed on the construct of interest. Unfortunately, little theory
exists to suggest that definitive time metric that is appropriate for studying most constructs in organizational contexts (Dansereau, Yammarino, & Kohles, 1999).

When short time intervals are examined (e.g., hours, days), they typically take the form of repeated measurements of an individual's standing on a construct used to assess the amount of transient error present in the measurement of a given construct (e.g., when assessing test-retest reliability; Feldt & Brennan, 1989). In such investigations, the units of the occasion-facet are typically hours, days, or months. The assumptions behind such an investigation are that (a) the individual’s true standing on the construct being assessed by the measure remains relatively constant over the short-term (long term stability is not typically a key issue in such investigations), and (b) variation in the individual’s standing on the construct across occasions reflects transient error variance (Schmidt et al., 2000).

Data collected over longer time intervals (e.g., years) are usually gathered in studies that examine individuals' personal development or learning, and assess the relative stability of their standing on a construct over an extended period of time (e.g., job satisfaction, Arvey, Bouchard, Segal, Abraham, 1989; job performance, Farrell & McDaniel, 2001; work values, Strong, 1955). In such investigations, the units of the occasion facets are typically months, years, or even decades. Furthermore, in such investigations, long-term occasion-related variation is not viewed as transient error (as is the case with short-term occasion-related variation), but rather as variance that is explainable through learning or developmental theory. As was the case with other loosely defined facets in the VAA, facets tapping short-term or long-term occasion-related variation may be included in any given construct’s variance architecture should data be available to provide stable estimates of their contributions.
Summary of VAA Facets

The above discussion described the facets that constitute the VAA. Decades of research in a variety of fields have indicated that variation related to these facets is present in many constructs of interest to organizational researchers. Table 2 provides a brief synopsis of the units of the facets of interest in the VAA.

Table 2. The Structural Content of a Construct's Variance Architecture

<table>
<thead>
<tr>
<th>Facet</th>
<th>Description of Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Culture</td>
<td>Cultures- countries, local regions- loosely defined&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Organization</td>
<td>Organizations</td>
</tr>
<tr>
<td>Group</td>
<td>Groups- workgroups, departments, business units- loosely defined</td>
</tr>
<tr>
<td>Job</td>
<td>Jobs, occupations, job families- loosely defined</td>
</tr>
<tr>
<td>Individual</td>
<td>Individuals (i.e., primary targets of measurement)</td>
</tr>
<tr>
<td>Item</td>
<td>Items or behaviors comprising a construct’s measures</td>
</tr>
<tr>
<td>Observer</td>
<td>Observers- raters or judges of an individual's standing on a construct</td>
</tr>
<tr>
<td>Stimulus</td>
<td>Exercises or events that evoke an individual’s standing on a construct</td>
</tr>
<tr>
<td>Occasion</td>
<td>Occasions- times at which the construct is measured (short-term / long-term)</td>
</tr>
<tr>
<td>Interactions</td>
<td>Each 2-way pairing and n-way grouping of the units of VAA facets</td>
</tr>
</tbody>
</table>

Note. “Loosely defined” indicates that the level at which units of this facet are defined may vary. Estimates of variation stemming from such facets will depend on the detail of the data available to describe a construct’s variance architecture.

Benefits of the Variance Architecture Approach

The present section elaborates on several of the benefits that the VAA can provide researchers in their study of organizational constructs. The benefits derive from three primary pieces of information that that will result from this approach. The first piece of information obtained from the VAA is point estimates of the absolute and relative amounts of total construct variance accounted for by each facet and interaction term contained in a given construct’s architecture. The second piece of information that results from the VAA is estimates of the stability of the contribution of each facet (and related interactions) to the total variance in a construct across different contexts (e.g., organizations, cultures, and time). The third piece of
information that will result from the VAA is an assessment of the degree to which the current research literature surrounding a construct allows for a full delineation of the construct's variance architecture. This latter information will arise as a by-product of the research process used to delineate variance architectures of constructs (detailed in later sections). Taking these three pieces of information as the primary output of any given VAA research effort, attention can now be turned to how obtaining this information can benefit organizational science and practice.

Identifying the Relative Contribution of Different Sources of Variance in Constructs

One premise behind the VAA is that organizational researchers and practitioners can benefit by focusing their efforts on prediction, explanation, and control of the target construct at levels across which it varies substantially. The VAA delineates the relative dispersion of variance in a construct across the facets of its variance architecture and thus identifies where the majority of variance in a given construct lies. Obtaining such information is beneficial because it can provide guidance as to where specific theory building, research, and intervention efforts may best focus.

For example, let’s say that relative to individual- and organization-level variation, VAA analyses reveal much of the variance in deviant behavior appears at the departmental-level within organizations. Such findings would indicate that devoting more attention to identifying and attempting to manipulate specific variables that account for the variance in deviance at the department-level may be the most salient way to reduce the incidence of deviant behavior in the workplace. Such department-level covariates of deviance might take the form of: (a) departmental rules, policies and procedures, (b) departmental compositions (e.g., percentage of employees with drug-related arrest records), or (c) traditional department-level variables (e.g., department morale, department cohesiveness). It is important to note that the use of the VAA in this example is not to pinpoint what these specific departmental-level covariates are, but rather to determine whether
the department-level facet is a relatively large source of construct variation worthy of consideration. Thus, the VAA offers researchers a mechanism for accelerating the development of the study surrounding an organizational construct by indicating potentially promising areas for future theory building and empirical research.

Establishing Upper Bounds of Predictability

In addition to identifying the most and least salient sources of variance in a construct, delineating a construct’s variance architecture can reveal the upper bound on the predictive ability of specific variables stemming from a particular facet. This upper bound may be expressed in terms of the percentage of construct variance for which predictors stemming from a particular facet may potentially account. For example, if the group-facet accounts for 10% of the variation in individuals' job satisfaction, then one knows that any single group-level predictor (e.g., group cohesiveness) or set of group-level predictors can, at most, account for 10% of the total variance in the job satisfaction construct (Lance & James, 1999).

A related benefit of obtaining VAA related information would arise when evaluating the percentage of facet-specific variance accounted for by a particular variable. For example, although a particular group-level predictor may account for only 2% of the between-person variance in a construct, it may account for 20% of the group-level variance. The difference between these two values lies in what is considered in the denominator for forming the proportion of variance accounted for index (e.g., individual- and group-level variance vs. solely group-level variance). Whereas the former value reflects a typical $R^2$ index, Lance and James (1999) refer the latter proportion of variance measure as $\nu^2$, and discuss in length the benefits of using such an index to describe the magnitude of relationships among variables in situations where one is dealing with multiple levels of analysis. Estimates of the denominator used in calculating $\nu^2$ values for any predictor at any level may be provided by the VAA.
Clarifying Appropriate Levels of Analysis for Facets with Loosely Defined Units

In addition to the benefits mentioned above, the VAA can also help clarify the level at which the units of various VAA facets are best defined for a given construct. For example, one might closely examine variation due to the job-facet for the general cognitive ability construct through inclusion of two job-facets in its variance architecture, one with its units defined at the job-level and the other at the occupation-level. Such analyses might reveal that general cognitive ability varies primarily as function of occupation (e.g., professors, factory workers), rather than jobs within occupations (e.g., professor of psychology, professor of mathematics), suggesting that the level at which to define the units of the job-facet in examinations of the VA underlying general cognitive ability should be based on occupations rather than job. On the other hand, such analyses might reveal that both job- and occupation-level variance is apparent in general cognitive ability, suggesting that including multiple job-facets would be appropriate for the studies of the VA of general cognitive ability.

The VAA may also help clarify the contributions of other facets with loosely defined units such as the group- and culture- facets. For example, one can more closely examine variation arising from the group-facet for a given construct by including two or more group-level facets in the delineation of its variance architecture (e.g., one with its units defined at both the workgroup-level, and the other at the department-level). By including multiple group-level facets in a variance architecture model, one is able to distinguish the relative contribution of workgroup-level variance and department-level variance. Similar models could be examined to differentiate the relative amount of variation across sub-cultures and national-cultures in attempting to clarify variation across the culture-facet. In any of the above cases, one is only limited by the data available to delineate such variance architecture models.
Identifying the Stability of Facet Contributions

In addition to estimating the relative contribution of each VAA facet to the variance in a given construct, the VAA also provides an estimate of the stability of those contributions across context and time. The examination of the stability of facet contributions is of great importance in that it speaks directly to the potency of the benefits outlined in the preceding paragraphs. For example, examining the stability of facet contributions lends insight into the degree to which (a) relative facet contributions are context dependent, (b) the upper bound for the variance accounted for by facets change as a function of context, and (c) the level or levels at which the loosely-defined VAA facets are best defined for a given construct vary across contexts. Given the importance of describing variance architectures in terms of both point-estimates of the contribution of its various facets, as well as the stability of those contributions, describing a construct’s variance architectures can be paralleled to reporting the results of a clinical-intervention study. For example, two pieces of information are critical in the reporting of clinical-intervention study results, the effect size of the intervention (analogous to the point-estimate of a facet’s contribution in the VAA) and an indicator of variation of the effectiveness of that intervention (analogous to the estimate of stability of facet contributions in the VAA).

Considering the stability of facet contributions raises the issue of what contexts we should be concerned with examining the stability our point-estimates of facet contributions across within in the VAA (McGuire, 1983). In the clinical–intervention studies the context of interest is generally individuals (i.e., the stability of the effectiveness of treatment across individuals). Within the VAA, the contexts across which facet contributions may vary are many. In this initial exposition of the VAA, three contexts are proposed: (a) organizations, (b) cultures, and (c) time. The choice of these contexts stems from the broad contexts I/O researchers have traditionally been interested in generalizing across. For example, the organizational level has been the focus of
validity generalization (e.g., Hunter & Schmidt, 1990), country and culture level has been the focus of cross-cultural research (e.g., Hofstede, 1980), and time has been the focus regarding questions of dynamic criteria (e.g., Farrell & McDaniel, 2001). Although the VAA does not limit examining the stability of facet contributions across only these facets (e.g. one may look at stability across job families or industries as well), only the three aforementioned contexts are discussed to simplify the issue of examining stability in this initial exposition of the VAA.

A problem that arises when considering the stability of VA facets across these contexts, are that these facets themselves correspond to facets of the VAA (i.e., organization, culture, and occasion). Thus, examining the stability of facet contributions across any of these contexts requires that the facet corresponding to the context in question be removed from the architecture being examined for stability. For example, if one wished to examine the stability of a variance architecture across organizations, one would need to exclude organization as a facet from architecture being examined. Similarly, if one wished to examine the stability of a variance architecture across cultures, one would need to exclude culture as a facet from architecture being examined. Details on estimating the stability of variance architectures across organizations, cultures, and times will be discussed in later sections that address the VAA research process.

Comparing Different Constructs

As mentioned at the beginning of this section, one premise on which the VAA is based is that organizational researchers and practitioners can benefit by focusing their efforts on prediction, explanation, and control of a target construct at the levels across which it varies most. A second premise on which the VAA is based is that understanding how organizational constructs relate to each other can be enhanced by delineating and comparing their underlying variance architectures (Cattell, 1966). Typically, the empirical similarity of constructs is conceived of in terms of correlations among two or more variables. Although the correlation between two
variables is a useful index of their similarity, it only provides a single point of comparison. A primary benefit of the VAA is that it provides multiple points of comparison across which two or more constructs can be compared. Specifically, the shared basis of common facets underlying the VAA make it a good mechanism for comparing two or more constructs by identifying the degree to which constructs are similar or differ with regard to the relative amounts of variance accounted for by each of their common facets. Thus, one can potentially compare and contrast various organizational constructs based on the dispersion (i.e., differential distribution across facets) of their total variance among these facets, and the stability of such dispersion across time and broader contexts of interest to researchers (e.g., different cultures or organizations).

Comparing constructs’ variance architectures for points of similarity and difference can provide insight into how constructs’ relate to each other. For example, examining the variance architectures of multiple constructs might reveal that some constructs that are similar on substantive levels (e.g., such as individuals’ perceptions of justice and job satisfaction) also share similar architectures. On the other hand, comparing variance architectures of similar constructs such as perceptions of justice and job satisfaction might reveal that they differ along several facets. Such findings would suggest potential sources of discriminant validity among the constructs (Campbell & Fiske, 1959). For example, if the relative contribution of the group-facet was much higher for the job-satisfaction construct than the perceived justice construct, it might suggest that reasons for reduced correlations between the constructs (i.e., an indication of discriminant validity) lie at the group-level (i.e., in this example group-level variables appear to be accounting for variance in job satisfaction, but not justice perceptions).

Investigating the architectures of multiple constructs can also potentially provide insight as to why two or more constructs that are believed to relate in lay theory, may not be exhibiting strong relationships empirically (e.g., job satisfaction and task performance, Muchinsky &
Iaffaldano, 1985; self vs. supervisor ratings of performance; Murphy & Cleveland, 1995). For example, variance architectures underlying job satisfaction and job performance may be quite different, suggesting that sources of variation primarily affecting one are different than sources of variation primarily affecting the other. Conversely, investigations into constructs’ architectures may also reveal instances where constructs that do not appear to have a relationship on a particular substantive level (e.g., the individual-level), share very similar variance architectures. As illustrated by the examples above, the VAA has the potential to offer researchers insight into new and unexpected areas of research relating the constructs that might be overlooked through a comparison underlying variance architectures.

Although the above discussion primarily focused on the benefits of comparing the architectures of two different constructs, similar benefits can also arise by comparing the variance architectures of two or more dimensions of a multidimensional construct. Many of the constructs of interest to organizational researchers are multidimensional in nature (e.g., job satisfaction, job performance, work values). As was the case with comparing architectures of different constructs, the value of the VAA for comparing architectures of different dimensions of the same construct is that it provides multiple points of comparison which can help identify sources of variation that may lead to different levels of discriminant validity (e.g., where facet contributions differ for the two dimensions). Establishing the discriminant validity of the dimensions of a multidimensional construct is often a crucial part of demonstrating a construct is indeed multidimensional rather than uni-dimensional. In later sections, I delineate a framework and methodology for systematically comparing variance architectures of stemming from either multiple constructs or multiple dimensions of a single multidimensional construct.
Comparing Different Methods of Measurement

Discussion of the benefits of the VAA to this point may have made it appear that there is only one specific variance architecture that can be used to describe the dispersion of variance underlying a single construct or dimension of a multidimensional construct. However, this is not the case. Often individuals’ standing on a construct can be assessed using a variety of methods. Based on the relative strengths and weaknesses of the method used to assess the construct, different methods might reveal slightly different variance architectures for that construct compared to other methods. As was the case with comparing the architectures of different constructs, comparing the architectures revealed by different measures of the same construct can provide valuable insight to organizational researchers.

For example, by examining architectures of a single construct, as revealed by multiple methods, one has a means of triangulating the "true" variance architecture underlying the construct (Sackett & Larson, 1990). The value of obtaining such information can be seen in the previous sections that described the value of delineating a construct’s VA. Another benefit of comparing variance architectures revealed by different measurement methods is that it can offer new insight or reinforce existing rationale for why different methods of measuring the same construct are similar or differ. For example, if one were to examine self-ratings, peer ratings, and supervisor ratings of individual job performance, differences in variance architectures revealed by these different methods may or may not be consistent with the relative strength and weaknesses of these rating methods identified in the performance appraisal literature (Murphy and Cleveland, 1995; Putka & Vancouver, 2000). Based on the performance appraisal literature, one might expect more variance to be accounted for by the individual-facet in architectures of job performance based on supervisor ratings as opposed to self-ratings (e.g., Murphy & Cleveland, 1995). Such a finding would be consistent with the belief that self-raters tend to inflate (thus
leading to less variance at the individual-level) their own ratings of performance due to social desirability concerns, thus leading to less variance in self-ratings across the individual-facet (e.g., Mitchell, Green, & Wood, 1981).

Examining the architectures of the same construct as revealed by different methods can also benefit researchers by potentially allowing them to assess how reliable a given method generally is for assessing a construct at a particular level of analysis. Specifically, under certain data sampling assumptions, the variance decomposition method employed in the VAA can provide variance component estimates that can be used to conduct a large-scale generalizability study (Brennan, 1983). Given G-theory assumptions hold, it is possible to use the variance component information generated within the VAA to calculate generalizability coefficients for the assessment of the construct at any particular level of analysis the researcher wishes (e.g., Cardinet, Tourneur, Allal, 1981). Of course examining reliability of a measurement of a construct at a particular level is dependent on the theoretical appropriateness of operationally defining that construct as a simple additive composite of individual-level data at the level one wishes to consider (e.g., group-level composites of individual-level data; Chan, 1998; Morgeson & Hofmann, 1999).

*Ancillary Benefits of the VAA*

Although the primary benefits of the VAA stem from the delineation of construct’s variance architectures, benefits arise from the VAA in the process of generating such estimates. For example, prior to estimating any parameter of a construct’s variance architecture, data must be gathered with regard to each of the facets. As later sections reveal, one needs a great deal of data across the units of the facets that constitute a construct's variance architecture in order to generate stable estimates of their contributions to the total variance in a construct. Unfortunately, large amounts of data on each facet may not always be readily available in the literature for a
construct (Cattell, 1966). Thus, in attempting to delineate a construct’s architecture, one may find (a) areas of the research literature where little information on particular facets is available, (b) insufficient coding of facet-related information (e.g., researchers failing to code for group membership when conducting individual-level research), or (c) data stems only from measurement designs or sets of organizations whose structuring prevents the unique estimation of facet effects and their interactions (e.g., fully nested data structures). As the present study will illustrate, the process of gathering data for delineating a construct’s variance architecture reveals shortcomings in the current literature. Thus, the process of attempting to delineate a construct’s variance architecture can reveal where systematic weaknesses (e.g., lack of data, coding, or confounding of facet effects) in the research literature’s ability to provide a complete description of the variance in a construct may lie. Identifying such areas can inform researchers where gaps in our knowledge regarding a construct may lie, and provide a diagnostic aide for potential future areas of research.

**Summary of VAA Benefits**

As discussed above, the VAA can provide many benefits to organizational researchers and can provide impetus for future research regarding a variety of constructs. Table 3 summarizes these benefits, and provides examples of research questions that may be addressed within the VAA with regard to each of the benefits listed. Although the VAA is not meant to be a solution to the person-situation debate in the organizational sciences, it does provide researchers a much needed systematic framework for conceptualizing and estimating the multiple sources of variation that underlie a construct. As Table 3 reveals, such a framework has much to offer organizational science.
Table 3. Benefits and Sample Research Questions Stemming from the VAA

<table>
<thead>
<tr>
<th>Benefits of the VAA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identifies areas where theory-building, prediction, and intervention efforts may have most / least impact.</td>
</tr>
<tr>
<td>- Where does the most / least variance in a given construct lie?</td>
</tr>
<tr>
<td>Establishes upper bounds on the predictability for variables stemming from each facet of variation</td>
</tr>
<tr>
<td>- What is the upper bound on the predictive ability of specific variables stemming from a particular facet of interest on a given focal construct?</td>
</tr>
<tr>
<td>Clarifies appropriate levels of analysis facets with loosely-defined units</td>
</tr>
<tr>
<td>- If a given construct exhibits variation across jobs, does that variation primarily arise from differences between specific jobs, occupations, or job families?</td>
</tr>
<tr>
<td>Identifies the degree of stability in the dispersion of variance across facets in a given construct.</td>
</tr>
<tr>
<td>- How stable are VAA facet contributions to variance in a construct across different organizations, cultures, and time?</td>
</tr>
<tr>
<td>Provides a unified framework for comparing different (a) constructs, (b) methods of measuring a single construct, and (c) dimensions of a multidimensional construct, across multiple levels of analysis.</td>
</tr>
<tr>
<td>- How do the architectures revealed by different methods of measuring the same construct compare?</td>
</tr>
<tr>
<td>- How do the architectures of different dimensions of the same construct compare?</td>
</tr>
<tr>
<td>- Do differences in variance architectures of constructs speak to potential sources of discriminant or convergent validity among them?</td>
</tr>
<tr>
<td>- Can some constructs be described much more simply than others (i.e., as a function of far fewer random facets than others)?</td>
</tr>
<tr>
<td>- Do the variance architectures underlying some sets of constructs differ in systematic ways from architectures underlying other sets of constructs? For example, do the architectures of cognitive constructs tend to systematically differ in some way from the architectures of affective constructs?</td>
</tr>
<tr>
<td>Identifies scant areas of the research literature for a given construct</td>
</tr>
<tr>
<td>- Where is the literature currently lacking in terms of its ability to provide data for a full delineation of a given construct's architecture?</td>
</tr>
<tr>
<td>- For what facets has little or no data been collected?</td>
</tr>
<tr>
<td>- Where is facet-related information not being recorded?</td>
</tr>
<tr>
<td>- Where is there a lack &quot;crossing&quot; in sampling or measurement designs that prevents the unique estimation of facet effects and their interactions?</td>
</tr>
</tbody>
</table>
The Variance Architecture Approach Research Process

The previous sections detailed the historical foundations, content, and benefits of the variance architecture approach. The focus of this section will detail the research process underlying the VAA. Specifically, I detail how constructs' variance architectures can be delineated via a research process consisting of three phases: (a) data gathering, (b) design specification, and (c) modeling of construct variance. The first of these steps is to gather the data that is available to delineate a construct's variance architecture from the research literature. At this first step, the large-scale nature of the VAA is realized as large amounts of archival data are gathered for subsequent analysis. The second step in the process is to lay out architectural designs on the data currently available that (a) most fully detail the construct's variance architecture, (b) allow the researcher to clearly address any a-priori research questions regarding the construct’s variance architecture, and (c) where possible, address any issues regarding the confounding of VAA facet effects in one’s data. The third step in the process consists of (a) modeling the variance in the construct in accordance with the architectural designs laid out in the previous step by fitting a General Linear Mixed Model (GLMM) to the data, and (b) assessing the stability of the dispersion of variance for the construct across contexts and time via meta-analytic methodology (e.g., Hunter & Schmidt, 1990).

Gathering the Data

One of the most salient features of the VAA is the data on which its analyses are based. Similar to traditional meta-analysis methods (e.g., Hunter & Schmidt, 1990), the data analyzed as part of a VAA effort are drawn from a diverse set of past studies to ensure their results generalize to the population of interest. Unlike traditional meta-analysis however, much more than summary statistics stemming from each study are needed in the VAA. Specifically, raw data on the construct of interest and units of the architectural facets sampled are required. The need for raw
data across a variety of facets substantially broadens the scope of the sampling effort involved in the VAA relative to meta-analysis.

For example, in the context of meta-analysis, if one wants to draw conclusions about whether the relationship between two constructs remains stable across *all* organizations in general, then one would sample correlations between the two constructs from studies that have been conducted across a wide variety of organizations. In meta-analytic studies in the organizational sciences, adequately sampling correlations from a variety of organizations (or whatever units the meta-analyst is concerned with generalizing across) is of prime concern to avoid second-order sampling error, just as adequately sampling individuals is in a single study to avoid traditional sampling error (Ashworth, Osburn, Callender, & Boyle, 1992; Hunter & Schmidt, 1990). Second-order-sampling error arises from not having an adequate sample of units on which the meta-analysis is being conducted (e.g., organizations) to generate stable estimates of the mean relationship of interest and the true level of variation observed among those estimates.

Unlike meta-analysis, because the goal of the VAA is to generate estimates of the contribution of *each* facet of a construct’s variance architecture to total construct variance, sampling across *all* facets in construct's variance architecture (to the extent possible) is of concern in the VAA (e.g., as opposed to only organizations). Further adding to the demanding data requirements is that, unlike meta-analysis, one is not simply sampling summary statistics (e.g., effect sizes) from a variety of studies, but rather raw observations or scores on the construct of interest, ideally going all the way down to the lowest unit (level) of specificity (i.e., the units of the highest order interaction term). Thus, to thoroughly delineate a variance architecture one must have a representative sampling of not only the population of the units at which the focal construct is typically operationalized (e.g., the individual-level), but also a representative sampling of the population of units of other facets in one’s design. In this regard, a nice parallel to the
multifaceted sampling requirements of the VAA appears in the context of Generalizability theory. In G-theory one must randomly sample across units of all facets of variation that contribute to the calculation of the G-coefficient (i.e., not only units of the object of measurement - typically individuals; Brennan, 1983).

As mentioned in the introduction, the probability that a researcher would obtain the breadth of sampling across each VAA facet in a single study is highly unlikely, thus the VAA relies on sampling of observations from past studies to accumulate data for delineating constructs variances architectures. Unfortunately, gathering even quantitative summary information (e.g., effect sizes of interest not reported in published articles), let alone raw data from past studies, is not a particularly easy task for organizational researchers to accomplish (Hunter & Schmidt, 1990). Given that raw data is needed, cooperation among researchers is vital to the successfully delineating a construct’s variance architecture. The reliance on cooperation of other researchers in this data gathering process will clearly be illustrated in the methods section of this study. In discussion ending the present study, alternative methods for more easily accumulating VAA-related data on constructs, as well as suggestions for facilitating cooperation among researchers to contribute data to VAA-related efforts are provided.

Given its epistemological similarity to meta-analysis, the VAA is potentially subject to many of the same methodological shortcomings of meta-analysis, in particular its reliance on judgment calls through out the process (Wanous, Sullivan, & Malinak, 1989). Wanous and colleagues (1989) outline a number of judgment calls that are made in the process of conducting a meta-analysis. Many of these judgment calls are pertinent in the gathering of data for both meta-analytic and VAA investigations. For example, judgment calls are involved in (a) defining the construct to be the target of a VAA investigation, (b) establishing criteria for including studies from which VA-related data will be drawn (e.g., published vs. unpublished, year of publication,
study quality) (c) deciding on how to conduct the search for studies that meet such criteria (e.g.,
computer vs. manual search) (d) selecting the final set of studies to be examined, and (e)
extracting VA-related data of interest from each study. Wanous and his colleagues (1989) suggest
that meta-analysts should carefully note the judgment calls they make in the process of gathering
data and conducting their meta-analysis. Given the similarity between data gathering methods of
meta-analysis and the VAA, details regarding the nature of the judgment calls made as part of the
current dissertation’s VAA-investigation into the OCB construct will be described in the Methods
section of this manuscript.

*Laying Out Architectural Designs*

Once data on a construct has been accumulated, the next step in the VAA research
process is to lay out several "architectural designs" on the variance in the construct based on the
data that is currently available in the applied literature. Laying out an architectural design on the
variance in a construct simply refers to specifying the VAA facets that will be included in the
delineation of a construct’s variance architecture and their orientation to one another (e.g.,
crossed vs. nested). This step is similar to the point in a Generalizability study when one specifies
a model for variance decomposition based on the measurement design underlying the construct
data. However, in the case of the VAA, neither the measurement design, nor the sampling designs
of the studies from which data is drawn is something that is determined a-priori by the VAA
researcher. Rather what facets can be included in a construct’s variance architecture as well as
their orientation to each other are dependent on the predominately used measurement and
sampling designs of the studies sampled. In light of the above observations, the primary focus of
this step of the VAA research process is on (a) assessing the structure and adequacy of the data
available for VAA analyses, (b) specifying architectures of varying level of detail based on the
structure of the data and a-priori research questions the researcher may have regarding the
construct, and (c) addressing issues arising from the confounding of facets in one’s architectural designs.

Assessing the Structure and Adequacy of Data for VAA Analyses

Assessing the structure and adequacy of the data available for VAA analysis involves determining the VAA facets for which data is available for the target construct and the general orientation of those facets to one another in the sampled data sets. Recall that an ancillary benefit of the VAA to organizational science is that the process of attempting to lay out architectural designs for a given construct allows one to identify areas of the research literature where little facet-related information is available for a given construct. At this step, the way the VAA achieves this benefit becomes clearer. Laying an architectural design on a construct’s variance forces one to systematically consider the VAA facets for which one has data and their general orientation to each other.

Ideally, architectural designs laid out for constructs would comprise all VAA facets, where each facet is fully crossed with the others (to avoid issues of confounding that arise from nested facets). An architectural design consisting of all VAA facets fully crossed with each other can be viewed as an ideal or conceptual model of construct variance. Such an ideal model is valuable to work from because it serves as a comprehensive map of the forms of variation that may underlie a construct (Cattell, 1966). Unfortunately, given the theoretical, practical and organizational constraints confronting organizational researchers, such an ideal is not likely to be achieved. As stated earlier, data available for VAA-related analyses may lack in a number of ways (e.g., data not being collected across some facets, facet information not being coded for in studies, confounding of facet effects due to nested data structures). In attempting to lay out architectural designs based the data currently available, this information will become readily available because the VAA modeling process (discussed below) requires the explicit specification
of the design on which the variance decomposition process for a construct will occur (i.e., what VAA facets are present in the design and their orientation to each other, e.g., crossed or nested). Thus, the actual models that are examined in the VAA approach will not likely contain all facets, but rather only those facets for which data are available.

**Laying Out Architectures in Varying Levels of Detail**

Once the structure of the data has been determined (i.e., facets for which data is available and orientation to one another), one can begin laying out designs for examining a construct's variance architecture. In laying out such designs, one should consider not only the highest degree of detail the data could provide with regard to that architecture, but also what simpler architectures may be beneficial to examine. More specifically, even if one has large amounts of data on all the VAA facets for a particular construct, it may still be beneficial to look at a less detailed, more general description of the variance in a construct in order to address specific a-priori questions in a less ambiguous manner. For example, given that a distinction commonly made in the multilevel literature is simply between individuals, groups, and organizations, one could examine an architecture that simply estimates the contribution of each of these three facets to the variance in a construct. To see if such a simple architecture is sufficient for describing the nature of variance underlying the given construct, one can subsequently look at more detailed architectures to see if a significant decrease in residual occurs. If the simpler architecture is sufficient, residual variance will not significantly decrease when more facets are added to the model. Comparing architectures of varying level of detail in such a manner can provide insight as to the complexity of the architecture that is necessary to describe the variance underlying a construct. As I discuss in later sections, one can statistically compare simpler and more complex architectures within a hypothesis-testing framework to see if the more parsimonious architecture sufficiently describes the variance in a given construct.
Another reason to consider laying out multiple designs for a construct’s variance architecture is to address problems that arise if the structuring (i.e., orientation of the VAA facets to one another) of the majority of the data gathered for delineating a construct’s VA is such that it does not allow for the unique estimation of the contribution of each facet. To illustrate the problems that confounding creates and potential methods for dealing with them within the context of the VAA, two scenarios are presented below. To keep things simple, only a limited number of facets are introduced in the scenarios below. Throughout the description of these scenarios, keep in mind that smaller subsamples of the data gathered in each scenario may be structured in a manner that does allow for unique estimation of facet effects that are confounded in designs laid out in the majority of the data.

**Scenario 1.** A researcher has gathered observers' ratings of individuals' engagement in Organizational Citizenship Behavior (OCB) from numerous past studies to examine the variance architecture underlying the OCB construct. The majority of data this researcher has is structured such that individuals are nested within workgroups and these workgroups happen to be managed by an observer who provides OCB ratings for all individuals within that workgroup (each group has a different observer).

In Scenario 1, there are a number of confounded sources of variation. For example, the group facet (g), and observer facet (o) are completely confounded. The researcher in Scenario 1 may choose to label the design based on such data as individuals-nested-within-groups (i:g) or as individuals-nested-within-observers (i:o). Although the contribution of the effects of groups and observers are completely inseparable, the researcher may have theoretical reason to expect that one source of variation (let’s say the group-facet) is primarily driving variation in the construct across the group/observer facet and thus may consider “i:g” a more appropriate label, being
careful to qualify his/her results with the fact that group-related variance may reflect observer-related variance as well. Another source of confounding arises among the individual facet (i) and individual-by-group/observer interaction (i x g/o). Because individuals are nested within the group/observer facet, it is impossible to uniquely estimate variance in OCB attributable to individuals or the individual-by-group interaction.

**Scenario 2.** A researcher has gathered observers’ ratings of individuals’ levels of conscientiousness (i.e., one of the “Big Five” personality factors) from numerous past studies to examine the variance architecture underlying the conscientiousness construct. The majority of data this researcher has is structured such that a different pair of observers rated each individual on their level of conscientiousness.

There are a number of confounded sources of variation in Scenario 2. For example, the observer facet (o), and individual-by-observer (i x o) are completely confounded. Because observers are nested within the individual facet (o:i) it is impossible to uniquely estimate variance in conscientiousness attributable to observers or the individual-by-observer interaction. Moreover, variance due to individuals and “observer-pairs” is completely confounded. This latter confounding may be less of a concern than the confounding of groups and observers presented in Scenario 1. Specifically, in this case there appears to be little theoretical reason to expect that observer-pair-variance contributes much variance to conscientiousness across individuals. Specifically, one can argue that the proportion of variance tapped by the individual-facet that is attributable to differences in observer-pairs is likely vary low because the individual rating idiosyncrasies of single observers within pairs will likely cancel each other out within pair (e.g., random pairings of lenient and severe raters).

**Resolving Confounding within the VAA.** To resolve the confounding in the two scenarios provided above, one must have subsample of data where the facets of interest are oriented in a
manner that would allow for separation of these effects (e.g., crossing of the confounded facets). For example, in Scenario 1 it would be possible to separate the group and observer effects if “pockets” of the observations gathered from past studies were structured such that there were multiple observers per workgroup who provided ratings of individuals within those workgroups: (i x o):g. Alternatively, the researcher may have a subsample of data where observers rated individuals in multiple workgroups such that individuals are nested within workgroups and workgroups are nested within observers (i.e., i:g:o). Although these designs resolve some confounds in Scenario 1 (separating group and observer effects), they create new confounds. An ideal subsample would consist of the same individuals participating in all workgroups and all observers rate all individuals in the context of each workgroup (i.e., a fully crossed design: i x g x o).

One potential problem with the above solution concerns how ecologically valid it is to apply inferences regarding variance estimates drawn from environments where facets of the VAA are crossed, to environments where such facets are nested. For example, perhaps there is something fundamentally different about organizations where individuals and group facets tend to be crossed (e.g., matrix organizations) compared to those organizations with traditional top-down hierarchy where such facets tend to be nested (Bedeian & Zammuto, 1991), and these differences make transporting variance estimates from one type of organization to another more difficult. One potential solution to this dilemma is to assess similar of the two environments on other characteristics. For example, in discussing methods for estimating the relative contribution of individual and individual-by-observer effects in nested ratings designs (e.g., individuals nested within observers), Hoyt (2000) suggests researchers can use meta-analytic estimates for the individual, and individual-by-observer effects from studies where these facets were crossed (and thus able to be uniquely estimated), given that similar circumstances surrounded the measurement
(e.g., similar amount of rater training, similar scale used to measure the construct, purpose of measurement, etc). Such similarity could help justify the transportability of variance estimates gleaned from studies in which facets are crossed, to provide unique estimates of those same facets for designs in which they are confounded (e.g., nested designs). Given the reality of how organizations are typically structured, the probability of finding substantial subsamples of data that have fully crossed facets is not very likely. Indeed, having a variety of subsamples with various degrees nesting is more realistic. In light of this observation, one may alternatively deal with issues of confounding by revisiting theory surrounding a construct. Specifically, theory can inform researchers regarding which confounds are most worrisome (e.g., variation across each confounded facet would be expected in theory) and which are less worrisome (e.g., variation in only one of the confounded facets, assuming a pair of facets are confounded, would be expected in theory).

For example in Scenario 2, say the researcher had a subsample of data where each pair of observers rated several individuals each: (i x o):p. Such a design would allow the researcher to separate the observer and individual-by-observer confound observed in the o:i design, and would allow the researcher to estimate the individual, observer, and individual-by-observer effects, as well as test the hypothesis of whether observer-pair variance accounts for a substantial proportion of individual-level variance. Nevertheless, confounds arise in the individual, observer and individual-by-observer effects because they are nested within observer pairs. However, in this case theory may strongly suggest that their would be little reason to believe that the individual-by-observer-pair interaction would account for any of the individual-level variance, the observer-by-observer-pair interaction would account for any of the observer-level variance, and the individual-by-observer-by-observer-pair interaction would account for any of the individual-by-observer variance. If theory is strong with regard to clearly eliminating some effects of the
design, than some sources of potential confounding may be less worrisome then confounds where little theory exists to speak to their magnitude, and little crossed data is available to uniquely estimate the facets are confounded.

**Summary**

Although one would ideally like to have a fully crossed design with a large amount of data on each facet, due to the traditional hierarchical structuring of organizations, the manner in which constructs are assessed, confounding of facet effects are likely to occur (e.g., due to nesting). The strategy suggested for dealing with confounding problems in the VAA is to estimate a construct's architecture in as much detail as the data will allow based on the structuring of the majority of one’s data, and then follow such estimation efforts with designs based on subsamples of data that can address various instances of confounding that occur in the larger designs (Hoyt, 2000). The information garnered from examining such smaller designs can be used to provide estimates for how the variances arising from confounded effects in the larger initial designs are distributed across their confounded components.

Although this is the general strategy I suggest, there may not be enough subsamples of data that are structured in a manner that allows one to uniquely estimate all sources of confounded variation in one’s larger variance architecture design. In this regard, theory may be used to ascertain whether remaining confounds are potentially worrisome. Where theory is not available, and confounds remain, care needs to be taken by the researcher in recognizing the limitations confounding puts on the conclusions they can draw regarding the contribution of confounded facets (e.g., in Scenario 1, variance stemming from the group/observer facet may reflect true differences between groups in terms of their engagement in OCB, as well as idiosyncratic rating differences between observers in terms of their use of the rating scale).
Modeling Variance via the General Linear Mixed Model

Upon gathering data and laying out architectural designs on the variance in a construct, the next step in the VAA research process is to estimate the construct’s variance architecture parameters. The primary purpose of this step is to provide estimates of (a) the relative and absolute amounts of construct variance accounted for by each facet and interaction that constitute the variance architecture, and (b) the stability of those values across context (e.g., organizational, cultural) and time. The method proposed for doing this is based on the General Linear Mixed Model (GLMM; Green, Marquis, Hershberger, Thompson, & McCollam, 1999; Littell et al., 1996). In the following sections, I will detail how the GLMM provides a basis for decomposing variance within the variance architecture approach.

An Overview of the GLMM

The GLMM is a generalized version of the general linear model (GLM). In matrix notation, the common general linear model is as follows:

$$ y = Xb + e $$

Where $y$ is an $n \times 1$ vector of observations on one’s construct of interest ($Y$), and $n$ is the number of observations on $Y$. $X$ is an $n \times p$ design matrix of fixed effects, and $b$ is the $p \times 1$ vector of those fixed effects, where $p$ is the number of fixed effects in the model. Finally, $e$ is the $n \times 1$ vector of residual random effects (i.e., effects arising from factors not explicitly included in the model). The GLMM expands upon this model to include non-residual random effects. Namely, the GLMM can be expressed as follows:

$$ y = Xb + Zu + e $$

The components of this equation are the same as (1) with the addition of $Z$ and $u$. $Z$ is the $n \times q$ design matrix of non-residual random effects and $u$ is the $q \times 1$ vector of those non-residual random effects, where $q$ is the number of non-residual random effects in the model.
Both residual and non-residual random effects stem from the levels or units sampled from a random factor in a given design. For example, if “individual” was considered a random factor in a GLMM, a random effect would be associated with each individual sampled from the population of individuals. Random factors are typically differentiated from fixed factors by the fact that the units of a random factor that are examined in a particular design represent only a random sample of the population of units in that factor (Jackson & Brashers, 1994). In the previous example, individuals are the units being sampled at random from some larger population of individuals. If “individual” were treated as a fixed factor, then the individual units contained in the model would be interpreted as the entire population of individuals rather than just a sample of them.

Random effects have historically been represented in linear models as deviations of a unit's score from the marginal mean of the random factor from which they stem (Brennan, 1983; Cronbach, Gleser, Nanda, & Rajaratnam, 1972; Hunter, 1968). Continuing with the example above, if “individual” was a random factor and the individuals sampled were the units of that random factor, the deviation of an individual's score from the expected value of all individuals' scores in the population would be the estimated random effect associated with that individual. Representations of random effects as deviations underlie most expositions of the random effects model underlying Generalizability theory, however such conceptualizations suffer from unnecessary restrictions that the sum of the effects must equal zero, which in practice may often not occur (DeShon, 1995). In more recent formulations of random effect models, random effects are viewed as realizations ( unknowable) of random variables with "probability properties" assigned to them (Searle et al., 1992). For example, probability properties typically assigned to random effects are that all random effects of a given random factor are independently and identically distributed (i.i.d.), and have an expected value of zero (mean) and a variance of $\sigma^2_{x}$, where the “x” subscript refers to the random factor in question (Searle et al., 1992). Following
from the assumption that such random effects are independently and identically distributed, the covariance between any two effects within a given random factor will be zero (Searle et al., 1992).

In random-effects models with multiple random factors, similar probability attributes are assumed for the distributions of random effects within each factor (i.e., for each random effect within a factor, i.i.d. \((0, \sigma^2_x)\)). Following from the assumption that random effects from each factor are independently sampled from their respective populations, the covariance between random effects between any two effects between factors will also be zero (Searle et al., 1992). Searle and his colleagues (1992) note that such probability properties typically assigned to random-effects in such models are just one set of properties that can be used and others may be specified within the GLMM that may better one’s data. Issues of which probability properties may best apply to random-effects models examined in the VAA will be addressed in later sections, but for current exposition of the GLMM these standard properties will be used. As Searle and his colleagues (1992) state, even these "elementary properties lead to enough difficulties insofar as estimation is concerned that alternatives seldom get used" (p. 10).

The estimate of the population variance of the random effects of a random factor (denoted as \(\sigma^2_x\) above) is called a variance component (Bryk & Raudenbush, 1992; Searle et al., 1992). These variance components provide the variance architectures of constructs. Specifically, each “facet” of construct's variance architecture will be considered a “random factor” contributing to the variance of the construct. The total variance in a construct can be decomposed into a set of variance components (one corresponding to each VAA facet and its interaction with other VAA facets). The sum of these variance components serves as an estimate of a construct’s total variance across units of all facets of the architectural design examined. Thus, each facet and
interaction term included in one's variance architecture can be viewed as accounting for a particular proportion of the total variance in the given construct of interest.

**Variance in the GLMM**

In matrix formulations of the GLMM, the variances and covariances of one’s observations on the focal construct of interest (let’s call this construct Y) are contained in V, the n x n covariance matrix of the n observations on Y. The formula for V for any architectural design is simply:

\[
V = ZDZ' + R
\]  

(3)

In this equation, Z is as stated above in (2), and Z' is its transpose. D is the q x q covariance matrix of non-residual random effects in vector u from (2). Finally, R is the n x n covariance matrix of the residual random effects in vector e from (2). Notice, there is no covariance matrix included for fixed effects in the formula for the variance in Y. This will always be the case no matter how many fixed effects there are in the particular GLMM one is studying. The reason for this is that such effects are not sources of random variation in Y; instead, they are fixed influences on mean levels of Y. The statistical literature recognizes this distinction in that variance across the fixed effects of a fixed factor is often referred to as a quadratic form (rather than a variance component; Searle et al., 1992).

The probability properties that are customarily employed in random-effects models serve to structure the D and R covariance matrices in a specific way (Searle et al., 1992). Specifically, under the random-effects model such matrices are constrained to have a “variance component structure” (Littell et al., 1996). A variance component structuring of these matrices means that all off-diagonal values are constrained to be 0. This constraint follows directly from the definition of the random-effects model laid down by Searle and his colleagues (1992), which indicates
covariances between the random effects in such models are assumed to be 0. Additionally, the
diagonal components of these matrices are variance components, and thus, by definition of the
random-effects model, are constrained to equality for any given effect within a random factor.
Specifically, the definition of the random-effects models is such that random effects within a
given random factor are independently and identically distributed with a common variance. In
later sections, I will address how the GLMM allows one to evaluate the robustness of these
common assumptions underlying the random-effects model for one’s data.

In any given architectural design examined, the primary parameters of interest are those
that constitute the diagonal elements of $V$. The parameters on this diagonal sum to the estimated
expected population variance of the observations on $Y$ across units of all facets of one’s
architectural design. Although, the covariances between random effects are constrained to be 0
under the random-effects model, this is not necessarily true for observations on $Y$. Specifically,
off-diagonal elements in the $V$ matrix that stem from the same facet can be nonzero, and their
covariance will be a function of the random factors with which they are associated. Such
characteristics of the GLMM will be further revealed in exposition of more complicated designs
introduced below.

The complexity of the process for decomposing variance in a construct primarily depends
on the number and nature of VAA facets included in the architecture examined. Generally, the
variance decomposition process becomes more complex as one adds more facets to one’s
architectural design. In the sections that follow, I will illustrate how the GLMM may be used to
estimate variance components for a variety of architectural designs. Although my primary
concern is with more complex architectures that describe the pattern of variation underlying a
given construct, for clarity's sake I will start my illustrations at the simplest level -- a design with
only one facet.
Examples of the Decomposition Process

Example 1: A single-facet design. Let’s consider a situation where a single observation on the focal construct of interest (Y) is taken on a random sample of individuals. Although modeling the variance in Y in this case is simply a matter of computing an estimate for $\sigma^2$ across individuals (e.g., $s^2$), for purposes of illustration I demonstrate this through use of the GLMM. As an example, let’s assume one randomly samples four individuals' responses to an organization-wide job satisfaction survey. In this case, "individual" can be considered a random factor, composed of one random effect for each individual sampled. When only one random factor is present in a design, the variance of the random effects stemming from it is typically modeled as residual variance (i.e., the effects are included in the e vector of residual effects and their covariance is modeled in the R matrix; Littell et al., 1996). Thus, for this simple design, the D and Z matrices drop out of (3) and the covariance matrix of the observations on Y can simply be expressed as:

$$V = R$$  

(4)

Based on the definition of the random-effects model, the random effects are independently and identically distributed, share a common variance, $\sigma_i^2$, and the covariances between effects are constrained to be zero. These constraints are reflected in Equation 5 below. Because individuals are the only source of variation in the design, the variance in Y is composed only of $\sigma_i^2$. Thus, the covariance matrix of observations on Y in this case is simply:

$$V = R = \begin{bmatrix}
\sigma_i^2 & 0 & 0 & 0 \\
0 & \sigma_i^2 & 0 & 0 \\
0 & 0 & \sigma_i^2 & 0 \\
0 & 0 & 0 & \sigma_i^2 \\
\end{bmatrix}$$  

(5)

Thus, the variance in Y is simply composed of a single variance component:

$$\text{Var}(Y) = \sigma_i^2$$  

(6)
In this simple case one can observe that the result of conceptualizing the decomposition of variance in Y through the lens of the GLMM provides the same result as simple intuition would suggest. The variance between individual observations is simply composed of one component, which could be easily estimated by calculating \( s^2 \). For more complex designs, partitioning the variance in Y becomes less straightforward.

Example 2: Individuals Nested within Groups. Expanding on the example above, let's say that the four individuals above were nested within two randomly sampled workgroups. In this case, the variance in Y may now arise from either the individual facet or the group facet. Formally, the equation for predicting observations on Y for the \( i_{th} \) individual in the \( j_{th} \) group can be expressed in terms of the GLMM as:

\[
\begin{bmatrix}
  y_{11} \\
  y_{21} \\
  y_{12} \\
  y_{22}
\end{bmatrix} = \begin{bmatrix}
  1 \\
  1 \\
  1 \\
  1
\end{bmatrix} \begin{bmatrix}
  \mu \\
  g_1 \\
  g_2 \\
  i_{11}
\end{bmatrix} + \begin{bmatrix}
  i_{12} \\
  i_{21} \\
  i_{22}
\end{bmatrix}
\]

(7)

As was the case in the first example, there is only one fixed effect in the equation and it is simply \( \mu \), the mean of all observations across individuals and groups (i.e., the grand mean). In this particular example, there are two random effects in the vector \( u \), \( g_1 \) representing the random effect of group 1, and \( g_2 \) representing the random effect of group 2. Written in non-matrix form, the equation for any \( y_{ij} \) is simply:

\[
y_{ij} = \mu + g_j + i_{ij}
\]

(8)

Where \( y_{ij} \) is the level of job satisfaction of the \( i_{th} \) individual in the \( j_{th} \) group, \( \mu \) is the estimated populations mean of all individuals, \( g_j \) is the random effect of group \( j \), and \( i_{ij} \) is random effect of the \( i_{th} \) individual in the \( j_{th} \) group.

In this design, there are two random facets, so the covariance matrix of Y will be a function of both D (the covariance matrix of non-residual random effects, i.e., group effects) and
R (the covariance matrix of residual random effects, i.e., individual effects) as specified in
Equation 3. Like Example 1, for modeling purposes the individual random effects will be
considered as residual effects and thus be modeled in R. By definition of the random-effects
model, we assume the group effects are i.i.d \((0, \sigma_g^2)\), with \(\text{cov}(g_i, g_j) = 0\) and the individual
effects are i.i.d \((0, \sigma_{i,g}^2)\), with \(\text{cov}(i_{ij}, i_{i'j'}) = 0\). Moreover, we assume the covariance between each
individual random effect and each group random effect is 0 (i.e., \(\text{cov}(g_j, i_{ij}) = 0\)). These
aforementioned constraints serve to structure the D and R matrices of the GLMM. Specifically, D
takes on the following structure:

\[
D = \begin{bmatrix}
\sigma_g^2 & 0 \\
0 & \sigma_g^2 \\
\end{bmatrix}
\]  \hspace{1cm} (9)

Based on the definition of the random-effects model for this second example, the R matrix takes
on the following structure:

\[
R = \begin{bmatrix}
\sigma_{i,g}^2 & 0 & 0 & 0 \\
0 & \sigma_{i,g}^2 & 0 & 0 \\
0 & 0 & \sigma_{i,g}^2 & 0 \\
0 & 0 & 0 & \sigma_{i,g}^2 \\
\end{bmatrix}
\]  \hspace{1cm} (10)

To generate a covariance matrix of observations for this example (V), one first takes the
D matrix and premultiplies it by \(Z\) (from Equation 7) and then postmultiplies it by \(Z'\) as stipulated
in Equation 3. This pre and postmultiplication results in the following matrix.

\[
ZDZ' = \begin{bmatrix}
\sigma_g^2 & \sigma_g^2 & 0 & 0 \\
\sigma_g^2 & \sigma_g^2 & 0 & 0 \\
0 & 0 & \sigma_g^2 & \sigma_g^2 \\
0 & 0 & \sigma_g^2 & \sigma_g^2 \\
\end{bmatrix}
\]  \hspace{1cm} (11)

To form the V matrix one then sums the resulting \(ZDZ'\) matrix (10) and the R matrix (11).
Taking this sum results in \(V\):
Thus, under constraints on D and R stemming from the definition of the random-effects model, the variance in Y is simply a sum of the variance component of the group effects ($\sigma_g^2$) plus the variance component of the individual effects ($\sigma_{ig}^2$), which is expressed as:

$$\text{Var} (Y) = \sigma_g^2 + \sigma_{ig}^2$$ (13)

There are several noteworthy observations regarding this decomposition of the variance in Y, particularly in comparison to the first example (single-facet design). For example, if one divides the variance component estimate for the group facet by the total expected variance for any observation on Y [i.e., $\sigma_g^2 / (\sigma_g^2 + \sigma_{ig}^2)$], one has an estimate of the total proportion of variance in observations ($y_{ij}$) that resides at the group-level. This quantity is commonly referred to as an intraclass correlation (e.g., Fisher, 1925; McGraw & Wong, 1996; Shrout & Fleiss, 1979). Along these same lines, notice that in the present design, the covariance between any two observations occurring in the same group is equal to the variance component for the group ($\sigma_g^2$), however observations from different groups are modeled as uncorrelated. This clearly demonstrates that although individual random effects are modeled as uncorrelated, this does not mean individuals' standings on the construct of interest are modeled as uncorrelated.

In contrast to V in this second design, the V for the single-facet did not allow for covariance between any of the observations on Y. Momentarily consider the implications of partitioning the variance in Y if the group facet was excluded from the model (as in Example 1), yet individuals were randomly sampled from groups. First, if groups did account for variation in Y, then the contributions of the individual facet would be grossly overestimated. In general, the
effects of leaving out a higher-level random facet from a design when it accounts for variance in
Y will tend to inflate the variance accounted for by lower-level facets nested within it
(Opdenakker & Van Damme, 1998). Second, observe that the covariance between observations as
a result of group membership would be ignored, thus the covariances among observations for the
model would not likely reflect reality (i.e., group membership likely leads to some degree of co-
variation (non-independence) among individuals' satisfaction within a particular group; Bliese,
2000; Hackman, 1992; Kenny & Judd, 1986; Mathieu & Kohler, 1990; Podsakoff, Ahearne, &

Unlike the first example, providing estimates for the variance components in a given
model becomes quite complicated when more than a single facet is included in one’s design (i.e.,
Var(Y) and its components can no longer simply be estimated by \( s^2 \)). Many methods for
estimating such variance components exist (e.g., Cronbach et al., 1972; Hartley, Rao, & Lamotte,
1978; Harville, 1977a; Searle et al., 1992). Furthermore, depending on the characteristics of the
data one is examining, some of these methods may provide better estimates than others. In light
of the complexity of these issues, as well as the fact that estimation considerations remain the
same for any given design, I will temporarily delay discussing estimators of variance components
until introducing other types of architectural designs.

Example 3: Individuals nested within groups and organizations. Now consider an
example where individuals are nested within groups and the groups they are sampled from are
nested within organizations. Assume that random sampling of the units of each of these three
facets (i.e., individuals, groups, and organizations) has occurred and thus the effects of any
individual i, any group j, or any organization k, can be considered random effects.

For purposes of illustration let's say that eight individuals are sampled from four groups
that are nested within two organizations. Two individuals are sampled from each group, and two
groups are sampled from each organization. The GLMM equation specifying \( y_{ijk} \), where \( y_{ijk} \) is the observation on \( y \) for the \( i^{th} \) individual nested in the \( j^{th} \) group in the \( k^{th} \) organization can be expressed as:

\[
\begin{bmatrix}
y_{111} \\
y_{211} \\
y_{121} \\
y_{221} \\
y_{112} \\
y_{212} \\
y_{122} \\
y_{222}
\end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ o_1 \\ g_1(o_1) \\ o_2 \\ g_1(o_2) \end{bmatrix} + \begin{bmatrix} i_{111} \\ i_{211} \\ i_{121} \\ i_{221} \\ i_{112} \\ i_{212} \\ i_{122} \\ i_{222} \end{bmatrix},
\]

(14)

Where \( o_1 \) and \( o_2 \) represent the random effects of organizations 1 and 2 respectively. The other four effects in the vector of non-residual random effects represent the effects of the \( j^{th} \) group nested in the \( k^{th} \) organization.

By definition of the random-effects model, the organization effects are i.i.d \((0, \sigma_o^2)\), with \( \text{cov}(o_1, o_2) = 0 \), the group effects are i.i.d \((0, \sigma_{g,o}^2)\), with \( \text{cov}(g_{jk}, g_{jk'}) = 0 \), and the individual effects are i.i.d \((0, \sigma_{i,g,o}^2)\), with \( \text{cov}(i_{ijk}, i_{ij'k'}) = 0 \). Moreover, by definition of the random-effects model, the covariance between each individual random effect, each group random effect, and each organization random effect is 0 (i.e., \( \text{cov}(g_{jk}, i_{ijk}) = 0; \text{cov}(o_k, g_{jk}) = 0; \text{cov}(o_k, i_{ijk}) = 0 \)). Based on the random-effects model definition for this three-level example, the D and R matrices of the GLMM take on the following forms\(^{13}\):

\[
D = \begin{bmatrix}
\sigma_o^2 & \sigma_{g,o}^2 & \sigma_{g,o}^2 & \sigma_{g,o}^2 \\
\sigma_{g,o}^2 & \sigma_o^2 & \sigma_{g,o}^2 & \sigma_{g,o}^2 \\
\sigma_{g,o}^2 & \sigma_{g,o}^2 & \sigma_o^2 & \sigma_{g,o}^2 \\
\sigma_{g,o}^2 & \sigma_{g,o}^2 & \sigma_{g,o}^2 & \sigma_o^2 \\
\end{bmatrix},
\]

(15)

\[
R = \sigma_{e,g,o}^2 I_8
\]

(16)
Where, $\sigma_o^2$ refers to the variance component of the organization effects, $\sigma_{go}^2$ refers to the variance component of the group effects, and $\sigma_{igo}^2$ refers to the variance components of the individual effects. Like the structuring of D and R in Example 2, observe that all off-diagonal elements are constrained to 0 (i.e., no covariances between random effects) and the variances associated with each effect within a factor are constrained to equality (i.e., the matrices have a variance component structure).

To complete the decomposition of variance in $Y$, one would premultiply $D$ by $Z$, postmultiply that result by $Z'$, and add $R$ as stipulated in (3). Hence, covariance matrix $V$ for this example is

$$V = \begin{bmatrix}
\sigma_o^2 + \sigma_{go}^2 + \sigma_{igo}^2 & \sigma_o^2 + \sigma_{go}^2 & \sigma_o^2 & \sigma_o^2 & \sigma_o^2 & \sigma_o^2 \\
\sigma_o^2 + \sigma_{go}^2 & \sigma_o^2 + \sigma_{go}^2 + \sigma_{igo}^2 & \sigma_o^2 & \sigma_o^2 & \sigma_o^2 & \sigma_o^2 \\
\sigma_o^2 & \sigma_o^2 & \sigma_o^2 + \sigma_{go}^2 + \sigma_{igo}^2 & \sigma_o^2 & \sigma_o^2 & \sigma_o^2 \\
\sigma_o^2 & \sigma_o^2 & \sigma_o^2 & \sigma_o^2 + \sigma_{go}^2 + \sigma_{igo}^2 & \sigma_o^2 & \sigma_o^2 \\
\sigma_o^2 & \sigma_o^2 & \sigma_o^2 & \sigma_o^2 & \sigma_o^2 + \sigma_{go}^2 + \sigma_{igo}^2 & \sigma_o^2 \\
\sigma_o^2 & \sigma_o^2 & \sigma_o^2 & \sigma_o^2 & \sigma_o^2 & \sigma_o^2 & \sigma_o^2
\end{bmatrix} \quad (17)$$

This model reveals that the variance in $Y$ is simply a sum of the variance components of the random facets contained in the model, namely:

$$\text{Var}(Y) = \sigma_o^2 + \sigma_{go}^2 + \sigma_{igo}^2 \quad (18)$$

There are a variety of noteworthy observations that can be made regarding the variance and covariance of observations on $Y$ in this example. First, the expected covariance between any two observations that do not stem from the same organization is zero. Secondly, the expected covariance between any two observations within an organization depends on whether those observations stem from individuals who are in the same group. If the observations stem from individuals in the same group, then the expected level of covariation is $\sigma_o^2 + \sigma_{go}^2$. If the
observations stem from individuals in different groups in the same organization, their expected level of covariation is $\sigma^2$. Perhaps more importantly, via Equation 18, one can easily identify the absolute and relative contributions of the organization, group, and individual facet to the estimated total construct variance of $Y$. These estimated contributions of each VAA facet to construct variance serves as the primary output of the VAA.

*Example 4: Adding a Crossed Random Facet.* The fourth example adds significantly more complexity to the previous designs by introducing the presence of a crossed, random facet, namely multiple items that tap the construct of interest are examined for each individual sampled. One of the characteristics of the designs discussed thus far is that they are all completely nested designs (Goldstein, 1995; Kirk, 1995). In the design illustrated in the present example, a behavior (item) facet is crossed with the three facets discussed in Example 3.

For purposes of illustrating the decomposition of variance for such a design, say that individuals' scores on each of two behaviors are examined. Further, assume that as in the last example, eight individuals were randomly sampled from four workgroups that were nested in two organizations. Assume also that the two behaviors examined represent a random sample of the population of behaviors that can be used to indicate the focal construct. Given such data, variance components stemming from four sets of random main effects can be estimated in this design, namely individuals, groups, organizations, and behaviors, as well as components stemming from the three interaction terms: organization-by-behavior, group-by-behavior, individual-by-behavior (completely confounded with the residual term).
For this example, the GLMM equation for the observations on $Y$ would be of the form:

$$
\begin{pmatrix}
  y_{1111} \\
  y_{1112} \\
  y_{2111} \\
  y_{2112} \\
  y_{1121} \\
  y_{1212} \\
  y_{2211} \\
  y_{2212} \\
  y_{1121} \\
  y_{1222} \\
  y_{2121} \\
  y_{2222}
\end{pmatrix} = \begin{pmatrix}
  1 \\
  1 \\
  1 \\
  1 \\
  1 \\
  1 \\
  1 \\
  1 \\
  1 \\
  1 \\
  1 \\
  1 \\
  1 \\
\end{pmatrix} \begin{pmatrix}
  \mu \\
  \beta_1 \\
  \beta_2 \\
  \gamma_1 \\
  \beta_3 \\
  \beta_4 \\
  \gamma_2 \\
  \beta_5 \\
  \beta_6 \\
  \gamma_3 \\
  \beta_7 \\
  \beta_8 \\
  \gamma_4 \\
\end{pmatrix} + \begin{pmatrix}
  e_{1111} \\
  e_{1112} \\
  e_{2111} \\
  e_{2112} \\
  e_{1211} \\
  e_{1212} \\
  e_{2211} \\
  e_{2212} \\
  e_{1121} \\
  e_{1222} \\
  e_{2121} \\
  e_{2222} \\
\end{pmatrix}
$$

(19)

In non-matrix form, Equation 19 for any given $y_{ijkl}$ can be expressed as:

$$
y_{ijkl} = \mu + \alpha_i + g_i(o_{ik}) + i_i(g_{ijkl}) + b_{kl} + e_{ijkl}
$$

(20)

Where $y_{ijkl}$ is the observed score of the $i^{th}$ individual nested in the $j^{th}$ group nested in the $k^{th}$ organization on the $l^{th}$ item. Once again working from the definition of the random-effects model: the organization effects are i.i.d $(0, \sigma_o^2)$, with cov $(\alpha_i, o_{ik}) = 0$, the group effects are i.i.d $(0, \sigma_g^2)$, with cov $(g_i, g_jk) = 0$, the individual effects are i.i.d $(0, \sigma_g^2)$, with cov $(i_i, i_{ijl}) = 0$, the behavior effects are i.i.d $(0, \sigma_b^2)$ with cov $(b_{kl}, b_{l'}) = 0$, the organization-by-behavior interaction effects are i.i.d $(0, \sigma_{ob}^2)$ with cov $(ob_{kl}, ob_{l'l'}) = 0$, the group-by-behavior effects are i.i.d $(0, \sigma_{go}^2)$ with cov $(g_i, g_jk) = 0$, the individual-by-behavior (residual) effects are i.i.d $(0, \sigma_{go}^2)$ with cov $(e_{ijkl}, e_{ijl'}) = 0$. Moreover, by definition of the random-effects model, the covariance
between each set of random effects is constrained to be 0 (e.g., \( \text{cov}(g_{jk}, i_{ij}) = 0; \text{cov}(o_k, g_{jk}) = 0; \text{cov}(o_k, i_{ij}) = 0 \)).

Based on the definition of the random-effects model, variance component structures are once again are imposed on the D and R matrices of the GLMM, such that:

\[
D = \begin{bmatrix}
\sigma_o^2 I_2 & & & \\
& \sigma_{g,o}^2 I_4 & & \\
& & \sigma_b^2 I_2 & \\
& & & \sigma_{ob}^2 I_4 \\
& & & & \sigma_{g:ob}^2 I_8 \\
& & & & & \sigma_{g:o}^2 I_8 \\
\end{bmatrix}
\]  \quad \text{(21)}

\[
R = \sigma_{g:o \times b}^2 I_{16}
\]  \quad \text{(22)}

Premultiplying the D matrix specified in (21) by Z (from Equation 19), postmultiplying it by Z', and adding R (from Equation 22), results in the covariance matrix of observations on Y for that is summarized by the following equations\textsuperscript{16}:

\[
\text{Var}(Y) = \sigma_o^2 + \sigma_{g,o}^2 + \sigma_{g:o}^2 + \sigma_b^2 + \sigma_{ob}^2 + \sigma_{g:o \times b}^2 + \sigma_{g:o \times b}^2
\]  \quad \text{(23)}

\[
\text{Cov}(Y_{ijkl}, Y_{i'j'kl'}) = \sigma_o^2
\]  \quad \text{(24)}

\[
\text{Cov}(Y_{ijkl}, Y_{ij'kl}) = \sigma_o^2 + \sigma_{g,o}^2
\]  \quad \text{(25)}

\[
\text{Cov}(Y_{ijkl}, Y_{ij'kl}) = \sigma_o^2 + \sigma_{g,o}^2 + \sigma_{g:o}^2
\]  \quad \text{(26)}

\[
\text{Cov}(Y_{ijkl}, Y_{ij'kl}) = \sigma_b^2
\]  \quad \text{(27)}

\[
\text{Cov}(Y_{ijkl}, Y_{ij'kl}) = \sigma_o^2 + \sigma_b^2 + \sigma_{ob}^2
\]  \quad \text{(28)}

\[
\text{Cov}(Y_{ijkl}, Y_{ij'kl}) = \sigma_o^2 + \sigma_{g,o}^2 + \sigma_b^2 + \sigma_{ob}^2 + \sigma_{g:o \times b}^2
\]  \quad \text{(29)}

\[
\text{Cov}(Y_{ijkl}, Y_{ij'kl}) = 0
\]  \quad \text{(30)}
Equation 23 indicates that the variance in Y is a function of seven variance components, namely, variance in Y due to individuals, groups, organizations, behaviors, the organization-by-behavior interaction, the group-by-behavior interaction, and the individual-by-behavior interaction, the last of which is completely confounded with residual variance.

The process involved in modeling the variance for the first four examples presented above serves as a sufficient background to understand the basic process involved in modeling the variance for the next few examples to be discussed. Because the next few examples can be viewed as extensions of the design above, their discussion will be limited to how they differ from Example 4.

**Example 5: Introducing Cross-Classification into a Design.** The fifth example adds to the previous design by adding a facet in which individuals may be nested separately in two or more facets. For example, individuals may be nested within both groups and jobs, yet neither jobs nor groups are necessarily nested within each other. In such a situation individuals are said to be cross-classified with respect to their group membership and the type of job they hold (e.g., Goldstein, 1994; Rabash & Goldstein, 1994; Raudenbush, 1993). In such designs, the variance across individuals represents the variance after removing group, organization, and job-related effects, whereas the other effects remain interpretable as indicated in earlier sections. In addition to this difference, the variance of Y would additionally be a function of a job facet and several job-related interactions, namely an organization-by-job, group-by-job, job-by-behavior, group-by-job-by-behavior, and organization-by-job-by-behavior interaction.

**Example 6: Accounting for Temporal Variation -- Including Occasion as a Facet.** As mentioned earlier, the occasion facet can be thought of in two ways. Namely, occasions may reflect a short-term time interval reflecting short-term fluctuation in terms of an individual’s standing on a construct as well as more of a long-term interval reflecting potential developmental
variation across the construct for a person or group overtime (e.g., Nesselroade, 1995). Including occasion, as a facet in a decomposition design is similar to including any of the above factors; however, including it will likely create several complications in terms of the decomposition of variance in the focal construct of interest.

First of all, the residual effects in the random-effect model, which have thus far been defined as uncorrelated, will likely now be correlated within individuals across time (Collins & Horn, 1995). The implications this has for the covariance among observations is it is likely that observations on individuals that are closer in time will be more highly correlated than those more distant in time will be less correlated (e.g., Ackerman, 1988; Edwards, 1991). To account for such "autocorrelation" among residuals within individuals, the GLMM allows for alternative structuring of the R matrix besides the structuring based on variance component constraints. For example, one can fit a first-order stationary autoregressive structure where correlations between residual effects become larger the closer any two observations come together in time (e.g., Bost, 1995). Alternatively, one may fit a simpler compound symmetrical structure to R that would constrain residual effects within individuals to some constant level of covariance (Jennrich & Schluchter, 1986; Wolfinger, 1993). Later sections will describe how models with alternative covariance structures for D and R can be compared via $\chi^2$ goodness-of-fit tests and relative fit induces to determine whether they provide a better fit to the data than the traditional variance components structure (e.g., Littell et al., 1996).

Given that the covariances of residual effects do not factor in to the equation for the variance in Y (recall Equation 3), one may wonder why they are of concern as the variance in Y is the primary focus of the VAA. Accounting for such correlated residuals is important in the context of the VAA because estimates of residual variance may be underestimated and in turn overestimates of the variance accounted for by other facets in the design may result if residuals
covary (e.g., Box, 1954; Bost, 1995; Maxwell, 1968; Rowley, 1989; Smith & Luecht, 1992). Previous research indicates that fitting more appropriate structures for \( R \) (e.g., compound symmetry, autoregressive) essentially removes the biasing effects that mis-specifying the model (e.g., resulting from imposing a variance component structure for \( R \)) may have on variance component estimates when correlated residuals are present in one’s data (Bost, 1995; Suen & Ary, 1989; Suen, Lee, & Owen, 1990).

**Summary.** The above examples provide an illustration of how the GLMM can be used to decompose variance under the definition of the random-effects model. Incorporating other facets of an organizational construct's architecture into such design would be a simple matter of extending these designs to include either more crossed or nested facets, or some combination of both. The examples above illustrate the parsimony that is possible for describing the variance in a given construct as a function of several general random sources of variation (i.e., facets of a construct's variance architecture), but also illustrate the flexibility inherent in the GLMM for examining other potential structures for the covariance matrices composing the GLMM besides those based on standard random-effect model constraints. Further discussion of this flexibility of the GLMM and model fitting issues will be discussed in later sections.

**Estimating Model Parameters**

The previous section illustrated how the GLMM could be used to describe the patterns of variation in a construct of interest as a function of one or more variance components corresponding to each of the VAA facets. What has yet to be discussed is how the parameters contained in the GLMM are estimated. Several potential methods for estimating variance components are available in the statistical literature and much research has been devoted to evaluating such techniques (e.g., Searle et al., 1992). For example, in two past reviews of the literature on variance components Sahai and his colleagues identified over 2700 articles regarding
the estimation of variance components published prior to 1983 alone (Sahai, 1979; Sahai & Kuhri, 1985). More recent work in the statistical literature has revealed that this large publishing trend with regard to issues surrounding variance component estimation has continued (e.g., Searle et al., 1992). Clearly much work has been done in this area, thus, for purposes of the present work, I will simply provide a brief review of some of the more common methods of variance component estimation, with an emphasis on their relative strengths and weakness for application within the VAA.

As alluded to in earlier sections, the choice of estimation procedure is determined for the most part by characteristics of the data with which one is dealing. Within the VAA, the ideal estimator would be able to retain its optimal statistical properties in the face of unbalanced designs due to the certainty that one would be dealing with unbalanced data when trying to decompose variance across organizational settings. Moreover, the ideal estimator would remain robust in the face of deviations from normality in distributions of random effects. Furthermore, the ideal estimator should retain its optimal statistical properties for a wide range of potential substantive-to-residual variance component ratios. Lastly, the ideal estimator within the VAA would be computationally efficient in that it could be generated even with the large amounts of data with which the VAA will most certainly encounter. Unfortunately, as reviewed below, no single variance component estimator currently available fulfills all of these ideal specifications. Nevertheless, some estimators that are available to researchers do appear more promising than others.

Estimators of Variance Components. In the sections that follow I briefly review three families of estimators that have served as the primary means of estimating variance components over the past three decades. The first set of estimators are based on the traditional ANOVA approach to variance decomposition. Within this set of methods are the expected mean square
(EMS) estimators which have been widely used in applications of generalizability theory (e.g., Brennan, 1983), as well as Henderson's (1953) extension of EMS methods to deal with unbalanced designs (Henderson, 1953). EMS estimators of variance components are calculated by computing the mean sums of squares for each term in a random-effects ANOVA model and equating each term to its expected value (e.g., Cronbach et al., 1972).

One particularly attractive attribute of EMS-based estimators is that they are non-iterative and require minimal computational power (Brennan, 1994). Furthermore, no normality assumptions with regard to the distribution of the random effects are made and, thus, they have been found to be relatively robust to departures of normality in terms of the distributions of the random effects (e.g., Searle et al., 1992). Additionally, EMS-based estimators provide unbiased, minimum variance, and efficient estimates of variance components when one is working with balanced designs with no missing data (Swallow & Monohan, 1984).

Although the above characteristics are desirable, EMS-based estimators have been severely criticized in the statistical literature because they can result in negative estimates for variance components, and lose their optimal statistical properties in designs that are unbalanced or have missing data (DeShon, 1995; Harville, 1977a; Searle et al., 1992). Given that the data being examined in the VAA will most certainly be quite unbalanced (e.g., due to variation in sizes of workgroups and organizations), EMS-based estimators are clearly not adequate for purposes of estimating variance components within the VAA.

Henderson (1953) proposed three methods that attempted to alleviate problems that traditional EMS estimators have in estimating variance components under unbalanced designs, which were also based on the ANOVA framework. Unfortunately, these methods appear to be derived more on grounds that they appeared to be reasonable solutions to the problem of estimating variance components within the ANOVA framework, not necessarily because they had
ideal statistical properties (DeShon, 1995; Searle et al., 1992). As such, these methods also tend
to be used infrequently to estimate variance components in the statistical literature.

**Minimum-Variance Estimators.** A second family of estimators that has been common in
the variance component literature are those known as minimum variance quadratic unbiased
estimators (MINQUE; Rao, 1971). MINQUE estimators require that starting values be specified
for each of the variance components in a given model, which are then used to produce minimum
variance estimates of variance components by minimizing a Euclidean norm through the solution
of linear equations (Searle et al., 1992). Although many varieties of MINQUE estimators exist
(based on different potential starting values for the variance components), one particularly
popular version of such estimators has been MIVQUE0 (e.g., Hartley et al., 1978). The
MIVQUE0 estimator is based on starting values of 1 for the residual variance component and 0
for all other substantive components. Beneficial qualities of MIVQUE0 estimators are that they
generally provide minimum variance, unbiased estimates of variance components, even in the
face of unbalanced designs (Rao, 1971; Hartley et al., 1978). Moreover, they require no normality
assumptions with regard to the distributions of random effects and do not require multiple
iterations to produce estimates.

Unfortunately, such estimates only retain their optimal statistical property of having
minimum variance at the starting values used and furthermore, different solutions will result if
different starting values are used, even when looking at the same set of data (Searle et al., 1992).
Moreover, the quality of MIVQUE0-based estimators of variance components has been found to
vary as a function of the ratio of magnitudes of substantive and residual variance components.
With MIVQUE0, estimates becoming quite poor when substantive components are greater than
the residual variance component (Swallow & Monohan, 1984). For these reasons, and the fact
that they can result in negative estimates, MIVQUE0 estimates do not appear to be a reasonable choice for adoption within the VAA.

Maximum Likelihood-Based Estimators. The third family of estimators are based on maximum likelihood estimation which has a long and respected history within the statistical literature dating back to its derivation by Fisher back in the 1920's (Fisher, 1925). Within this family of maximum likelihood estimators, two specific estimators are commonly employed to estimate variance components, namely full maximum likelihood (ML; Hartley & Rao, 1967), and the restricted maximum likelihood estimators (REML; Patterson and Thompson, 1971).

ML and REML estimators of variance components have a number of ideal characteristics. First, both retain their optimal statistical properties with both balanced and unbalanced designs, as well as when data are missing (Hartley & Rao 1967; Patterson & Thompson, 1971; Searle et al., 1992). Second, variance components estimates based on these methods are optimized over a parameter space that constrains the estimates to be non-negative in value (Searle et al., 1992). Third, although maximum likelihood estimators have been criticized based on their normality assumptions (with regard to the distribution of random effects), several authors argued and demonstrated robustness of REML estimators to mild and moderate violations of normality assumptions (e.g., Harville, 1977a; Marcoulides, 1990). Typically, REML is viewed as a slightly more favorable estimator compared to ML due to (a) its ability to also examine designs which include fixed effects, (b) its similarity to ANOVA-based estimators for solutions when data are balanced, and (c) its avoidance of "ridiculous" estimates of variance components that are sometimes produced with the ML method (Harville, 1977b, p.339; Searle et al. 1992). As Searle et al. (1992) concluded, such estimators are coming to be the preferred method of estimating variance components when one has unbalanced data.
Given these beneficial aspects of maximum likelihood based estimators of variance components and REML estimators in particular, the REML estimators of variance components currently appear to be the best choice for use in the VAA. As such, I will briefly explain how REML-based variance components estimates are produced. Describing REML estimators in more detail will help provide grounds for the discussion of the one major drawback of using REML based estimators (i.e., their lack of computational efficiency in the face of large data sets).

**REML Estimates of Variance Components.** Calculating variance components within the GLMM based on REML estimation can be explained by returning to the matrix formulation of the GLMM described earlier (see Equation 2). Because the random-effects model will be employed within the VAA, the fixed part of that equation \((Xb)\), can be reduced to simply, \(\mu\) (i.e., the grand mean of observations on \(Y\)). Thus, Equation 2 can be restated as:

\[
y = \mu + Zu + e
\]  

(31)

Based on Searle et al.'s (1992) exposition of REML variance component estimators, it will be helpful to re-express the nonrandom residual component of the model as:

\[
Zu = \sum_{x=1}^{r} Z_x u_x = \begin{bmatrix} u_1 \\ \vdots \\ u_r \end{bmatrix} \]  

(32)

Where the \(r\) subscript refers to the total number of non-residual variance components being estimated in a given design and the \(x\) subscript refers to the \(x^{th}\) non-residual random factor or interaction between random factors in the design. The vector of \(u_x\)'s is composed of all the random effects in the model distinguished by the particular random factor or interaction they stem from (indicated by the \(x\) subscript). The row vector of \(Z_x\) matrices indicates the \(n \times q_x\) partition of the design matrix \(Z\) corresponding to the \(x^{th}\) set of random effects.

Furthermore, one can re-express the vector of residual effects in Equations 2 and 31 as:
\[
\mathbf{e} = \mathbf{I}_n \mathbf{u}_0 = \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix} = \mathbf{Z}_0 \mathbf{u}_0
\]  
(33)

In this equation, \( \mathbf{I}_n \) is an identity matrix of magnitude \( n \times n \) (recall \( n \) is the number of observations on \( \mathbf{Y} \)) which one can name \( \mathbf{Z}_0 \), and \( \mathbf{u}_0 \) is simply the vector of residual random effects. In light of this re-expression of the residual vector, Equation 31 can be further reduced to:

\[
y = \mu + \sum_{x=0}^{r} \mathbf{Z}_x \mathbf{u}_i
\]  
(34)

Note, that the starting index on the summation sign is now 0, indicating that the residual vector is now being summed into the model. Given this new notation, the variance-covariance matrix of observations on \( \mathbf{Y} \) from Equation 3 can be re-expressed as:

\[
\mathbf{V} = \sum_{x=0}^{r} \mathbf{Z}_x \mathbf{Z}_x' \sigma^2
\]  
(35)

In this equation, \( \mathbf{Z}_x \) is as stated above, \( \mathbf{Z}_x' \) is simply the transpose of \( \mathbf{Z}_x \), and \( \sigma^2 \) simply refers to the variance component of the \( x^{th} \) random factor or interaction term.

To estimate the \( r + 1 \) variance components contained in Equation 35, one would work with the likelihood function of the vector of observations on \( \mathbf{y} \), namely:

\[
L = L(\mu, \mathbf{V} | \mathbf{y}) = \frac{e^{-1/2(y-\mu)'V^{-1}(y-\mu)}}{2\pi^{n/2}V^{1/2}}
\]  
(36)

Generating REML estimates for the variance components is a matter of finding the set of values for \( \mu \) and \( \mathbf{V} \) that maximize this function. In attempting to maximize this function, it is easier to work with its logarithm, which is more mathematically tractable (Searle et al., 1992). The logarithm of this likelihood function (denoted with a lower case \( L \)) is:
REML-based estimates of the variance components for a given design maximize the fit of the log-likelihood function (Equation 37) to the residual $y$ values after partialling out the fixed effects (in the case of a random-effects model, $\mu$). To maximize this function, one would (a) differentiate the partialled log-likelihood function with respect to the variance components in $V$, which results in:

$$
I_{\sigma^2} = \frac{\partial I}{\partial \sigma^2} = -1/2 \text{tr}(V^{-1}Z_xZ'_x) + 1/2 y'V^{-1}Z_xZ'_xV^{-1}y
$$

(38)

then (b) equate this partial derivative to zero, and (c) solve the resulting set of highly complex nonlinear equations (DeShon, 1995, Searle et al., 1992).

Given the complexity of the resulting nonlinear equations, iterative numerical techniques are required to solve them. Although several potential algorithms are currently available in the statistical literature for solving such equations (e.g., Newton-Raphson, Marquardt, Fisher scoring; Quasi-Newton, EM; see Searle et al., 1992 for a comprehensive review), a combination of two such algorithms appears to currently be in favor in the statistical literature. Namely, the common recommendation is to use Fisher scoring methods to work through the first few iterations and then switch to the Newton-Raphson algorithm to complete the process (e.g., Jennrich & Sampson, 1976; Littell et al., 1996; Searle et al., 1992). The reason for the combination is that Fisher's scoring method appears to be more robust to poor starting values than the Newton-Raphson algorithm, yet the Newton-Raphson will typically result in a faster rate of convergence (Searle et al., 1992).

Although REML estimators have a number of desirable properties, generating REML estimates of GLMM parameters is an extremely computationally intensive procedure (Littell et al., 1996; Searle et al., 1992; SAS Institute Inc., 1999). For small amounts of data generating
REML-based estimates of variance components can be achieved via commonly available commercial software programs (e.g., SAS). However, when one wishes to estimate GLMM parameters in large data sets, the computational demands of the REML estimation method become more problematic. Given the large amount of data the VAA will focus on, software and hardware issues are of concern. More details on this issue, methods to work around estimation problems, as well as suggested software and hardware to use when generating estimates for GLMM parameters are addressed in Appendix A.

Summary. Based on the current status of the statistical literature, REML estimators are viewed as the most promising methods currently available to estimate variance components within the VAA (Searle et al., 1992). Until less computationally intensive methods (e.g., RAVE estimators; Mehrotra, 1997; discussed in Appendix A) are more fully developed for the complex designs likely to be confronted in the VAA, attempting to work around the computational limitations imposed by algorithms that produce REML based estimates appear to be the best strategy to adopt for purposes of the VAA.

Model Fitting Issues in the VAA

Having discussed estimation of GLMM parameters, attention is now turned back to the issues of fitting models that best describe the variance architecture of a given construct. Three primary issues with regard to model fitting arise within the VAA. First, because all architectural designs within the VAA contain only random effects (with the exception of the single fixed intercept), all of the designs can be called completely random. Given this fact, models of variance that specify more general architectural designs can be compared directly to models with more detailed architectural designs (in terms of the number of VAA facets included). Thus, one issue of model fitting that must be addressed within the VAA is the level of detail necessary to adequately describe a construct's variance architecture. Perhaps for some constructs, a very
simple, general architecture may suffice, whereas for others, a more complex, detailed architecture may be needed.

A second aspect of model fitting that needs consideration within the VAA concerns examining the fit of different models of the variation within a specific architectural design, namely examining alternative specifications for the structures of the D and R covariance matrices for a given architectural design. In this respect, fitting a GLMM to the data for a given design is an exercise in modeling the covariance structures of D and R (Jennrich & Schluchter, 1986; Littell et al., 1996; Wolfinger, 1993). As alluded to earlier, in some situations the variance component structurings of D and R that commonly underlie the random-effects model may not provide the best descriptions of the variance underlying some designs and/or constructs. As I argue below, however, it is likely that the variance component model is sufficient, except perhaps in architectures where occasion is included as a facet (as described earlier).

Lastly, a third aspect of model fitting that needs to be considered is the degree to which contributions of facets identified in a given architectural design remain stable across context and time. This last model fitting issue addresses the question of whether a construct’s variance architecture is best conceptualized as a relatively stable entity or an entity that is better described in terms of a set of mean facet with corresponding variabilities reflecting variation in those facet contributions across context or time. Perhaps some constructs’ architectures may be more stable than others and their relative stability may further depend on whether one examines stability across organizations, cultures, or time. In the sections that follow, methods for addressing each of the three model-fitting issues introduced above are examined.

Simple versus Complex Architectures - Number of VAA Facets. Given that all effects modeled in any VAA design are random (with the exception of the intercept), the relative fit of one random-effects model to the data compared to another can be evaluated statistically by
comparing the models' maximized log-likelihood values (assuming that they are both based on
the same set of data). Specifically, the difference between the two maximized log-likelihood
values of the models of variance that only differ in that one is composed of more random facets
than the other is $\chi^2$ distributed. The degrees of freedom for such a $\chi^2$ is equal to the difference in
the number of parameters being estimated by each model (Littell et al., 1996). Thus, by
comparing the log-likelihoods and testing the significance of their difference based on a $\chi^2$-
distribution, one can provide an inferential test of whether the more complex architecture
provides a significantly better description of the variance in the construct compared to the simpler
model. Alternatively, one could also compare the variance component associated with the
residual term in each model. If the residual term accounts for less of variance in one model
compared to another, it is an indicator that the model with the lower residual variance accounts
for more substantive variation in the construct.

Although such a statistical fit test or direct comparison of the residual values are
important criteria to consider when comparing the adequacy of simpler versus more complex
architectural designs, there is also another criterion to consider, namely, parsimony. A more
complex design or model will almost certainly provide a better fit to the data. However, a more
complex design will also require the estimation of more parameters, and thus be a less
parsimonious description of the variance in $Y$. What is needed is an index that takes into account
the number of parameters estimated when determining the relative quality of one design
compared to another. Such relative fit indexes are readily available in the structural equation
model literature (e.g., Bentler, 1990). Relative fit indexes provide estimates of the improvement
of fit of a given model, relative to another model, and "penalize" the less parsimonious model
(e.g., Aikike's Information Criterion (AIC) or Schwarz's Bayesian Criterion (SBC); Littell et al.,
1996). Such relative fit indexes arising from complex and simple architectural designs on the
variance in a given organizational construct can be directly compared to aide in the judgment of which model provides a better fit to the data (assuming once again both relative fit indexes for each model were calculated on the same set of data; Littell et al., 1996).

Alternative Covariance Specifications for D and R. As stated earlier, the random-effects model that underlies the examples of the decomposition process within the GLMM described earlier is based on common probability properties that are assigned to the random effects (Searle et al., 1992). However, alternative probability properties may better reflect one’s data and, thus, covariance structures which impose different patterns of constraints other than those specified by the variance component definitions of the traditional random-effects model may be more appropriate.

Ideally, one should take a theoretical approach to specifying potential alternative structures for the D and R matrices. Based on theory surrounding the construct, one should determine what covariance structures are likely to describe the pattern of variance with the construct as a function of the random facets examined and compare such structures to a baseline structure for comparison. For example, one strategy is to always start by fitting a model with a variance component structure for D and R to serve as a baseline against which to compare other potential models. The reason why the variance component model would serve as a good baseline model is that it a) is relatively simple compared to other structures, b) it is the standard structure that underlies the random-effects model, and c) it ensures that the variance in Y is just a function of variance components of random facets and not covariances between them (addressed below). The latter would occur if random effects from different facets were correlated (Searle et al., 1992). Upon establishing such a "baseline" model of variance for a given architectural design one could compare the fit of this baseline model to a model that specifies another covariance structure for either D or R (doing both simultaneously, poses several estimation problems; Littell et al.,
1996) the researcher feels is more appropriate based on theory surrounding the construct in question or the facets examined (e.g., a compound symmetrical or autoregressive structure for R in designs including an occasion facet).

One can compare models with alternative covariance structures specified for D or R within a given design based on the log-likelihoods that result fitting different covariance structures to the data, as well as the relative fit indices associated with each alternative model (e.g., Singer, 1998). Thus, model fitting in this regard is essentially a process of comparing theoretically derived covariance matrices for D and R to determine which structuring provides the best fit to the data. Unlike comparing designs of varying detail however, the number of parameters that need to be estimated under alternative structures of the D or R variance-covariance matrices can exponentially increase with increases in the amount of units per facet that are sampled. Thus, relative to comparing architectures of differing levels of detail (in terms of number of facets included), the process of estimating models for different structures of D and R within the VAA and comparing them may be much more computationally intensive. Given that the best methods for currently estimating parameters of GLMM that underlie the VAA are iterative in nature and require the storage and inversion of matrices of extremely large magnitudes, the more parameters being estimated the heavier this already heavy computational burden would become (see Appendix A for examples).

In addition to being computationally burdensome, finding that alternative covariance structures for D and R provide a better fit to one's data can create further complications within the VAA. Namely, if alternative structures of the D covariance matrix provide a much better fit to one's data relative to a variance component structuring, a substantial loss of parsimony in describing the VA of a construct can be expected. Specifically, if alternative covariance structures for D allow the off-diagonal elements to be nonzero (i.e., the random effects are correlated), the
description of the variance in Y will not simply be the sum of variance components, but rather a sum of variance components and covariances. Thus, if such covariation among random effects reflected the reality of one's data, one's description of the variance architecture underlying Y would be quite complex.

It is important to note that although nonzero, off-diagonal elements in D may have a great impact on the parsimony with which the variance in Y can be decomposed, nonzero, off-diagonal elements in the R will not impact the description of the variance in Y. Specifically, even with covariances among residuals modeled in R, the variance in Y is still purely a function of variance components. Nevertheless, if such covariance is present in the data, it is important to model it due to the effects that not allowing nonzero covariances to vary may have on the accuracy of the estimated variance components (Bost, 1995).

Fortunately, it may be the case that alternative structures do not arise that often within the VAA. With the exception of models that have occasion as a facet, the variance components structuring of the D and R matrices should fit quite well in most cases. Specifically, there is no theoretical case for why one would expect covariances between the random-effects examined in the VAA, particularly given that they are independently and randomly sampled by several different past studies. Moreover, the variance component model has been shown to underlie similar types of data collected for decades in the generalizability theory literature (Shavelson, Webb, & Rowley, 1989). Given that the content of VAA does not differ substantially from what has been examined in the G-theory literature, it would appear to be difficult to generate a strong theoretical argument for a standard structuring for the D and R covariance matrices other than a variance component structuring, with the exception of a few isolated cases (e.g., designs with occasion as a facet).
Examining the Stability of Variance Architectures. Besides fitting alternative standard covariance structures for D or R (e.g., compound symmetrical, autoregressive), one may free specific constraints imposed by the variance component structuring in a manner that is consistent with what theory may imply. For example, one such constraint that may be relaxed is that of homogeneity of variance among units within nested facets (e.g., individual-level variance is identical in each organization examined; $\sigma_{i.o1}^2 = \sigma_{i.o2}^2 = \sigma_{i.oK}^2$). Examining potential heterogeneity speaks directly to the issue of the stability of estimates of architectural parameters across different contexts. Such heterogeneity may arise for several reasons. For example, Johns (1991) argued that organizations vary in terms of the strength of their HR systems for constraining variation in individual-level work behavior and performance among their members, such differential constraint may lead to variation between organizations in terms of the amount of variance between individuals on a given construct. Exacerbating the potential for such differences, variance component estimates for any given facet in a construct's architecture across organizations may also vary due to the relative amount of constraint imposed by different norms and values espoused by various organizations (e.g., Katz & Kahn, 1978; O'Reilly & Caldwell, 1985; Schneider, 1990). In light of the potential for such informal means of constraint, there may also be a case for heterogeneity stemming from variation in constraining mechanisms operating at higher levels as well (e.g., cultural-level constraints, Chao, 2000). Thus, when evaluating constructs' variance architectures, testing for the possibly of heterogeneity in facet contributions across units of such broad contextual facets is suggested.

Although testing whether such heterogeneity exists across any facets included in a construct's variance architecture as a function of context or time is possible via the GLMM (by freeing equality constraints imposed on variance component values D and R), there are two reasons why it would not be ideal to do so. First, as suggested above, fitting alternative
covariance structures for D and R would create an extremely great computational burden. Specifically, fitting such heterogeneous variance models would substantially increase the number of parameters that would need to be estimated within a model because a separate variance component estimate for each of the VAA facets in the design (and their interactions) other than the contextual factor (e.g., organization or culture) would have to be estimated for each unit of the contextual factor being examined as a source of heterogeneity. In light of these problems, and given the importance of determining whether a construct's variance architecture remains stable across these contexts, this particular aspect of model fitting can actually be recast in terms of meta-analyzing VAA facet contributions across units of the contextual factor of interest. This methodology (which I discuss below) is not only in line with the goals of the VAA, but also much less computationally burdensome than working entirely through the GLMM framework.

*Meta-Analytically Assessing Architecture Stability*

Assessing the stability of a variance architecture can be viewed as highly similar to examining models with heterogeneous variance structures with the GLMM alluded to above. However, there is a key difference—relying solely on the GLMM only provides for a test of model with heterogeneous variance components versus equality of variance components. Thus, the comparison made is in terms of absolute values of variance component estimates across units of the contextual factor. The meta-analytic methodology proposed here takes a slightly different approach in that it compares the relative contribution of a given VAA facet to construct variance across units of the contextual factor.

Before proceeding, understanding the difference between absolute and relative facet contributions is important to grasp. Consider, for example, the variance accounted for in job satisfaction by individuals in one organization on is .2, and in another, it is .4. At first glance, such values would appear to indicate that relative to other facets in the construct's architecture,
the contributions of the individual facet appear to be much greater in the second organization. However, if the total construct variance in those organizations were .8 and 1.60, then the relative contributions of the individual facet would be identical. Comparing relative values, in essence, standardizes variance component values based on the total construct variance within the given context (i.e., unit of the contextual factor, a specific organization, culture or time period). Although ideally it may be beneficial to examine the stability of facet contributions to total variance in both absolute and relative terms, the meta-analytic methods only exist for examining the stability of relative facet contributions, not absolute variance components.18

To provide a test of the stability of relative VAA facet contributions across units of a given contextual factor (e.g., organization, culture, time), one can use a combination of both the GLMM and meta-analytic methods to provide a solution that greatly reduces one's computational burden. Specifically, estimating the stability of VAA facet contributions can be achieved in three steps. First, estimate the variance components corresponding to the facets of a construct's VA by fitting a GLMM based on the construct’s VA for each unit of the contextual factor separately using variance component structures for D and R. Second, calculate the resulting relative contribution of each VAA facet to the total variance for each unit of the contextual factor (i.e., divide the variance component of each VAA facet within each unit of the contextual factor by the sum of all variance components calculated within the given unit of the contextual factor). Third, use meta-analytic homogeneity tests to determine whether the relative contribution of each VAA facet varies across units of the contextual factor after accounting for artifactual variation in the individual estimates that is due to sampling error (e.g., Q tests of homogeneity, Hunter & Schmidt's 75% rule).

As with most meta-analytic methods and related homogeneity tests, assumptions regarding normality in the distribution of the effect sizes of interest are often made. As such, a
normalizing transformation of the proportions of variance accounted for by each VAA facet within each unit of the contextual factor is recommended prior to carrying out the meta-analytic tests of homogeneity. As cited earlier, the proportions of variance accounted for by a VAA facet can also be viewed as intraclass correlations (Hocking, 1990; Hoyt & Kerns, 1999). Although ICCs are not typically normally distributed, normalizing transformations do exist for ICCs and, as an added benefit of implementing such transformations, standard errors for such normalized values are generated as well (Fisher, 1925; McGraw & Wong, 1996). Formulas for carrying out these normalizing transformations on estimated facet contributions are presented in Appendix B.

Armed with these normalized facet contributions and their associated standard errors one can employ Hunter and Schmidt's (1990) meta-analytic techniques to estimate the raw variance in proportions of variance accounted for by a given facet across units of the context versus sampling error. Formulas for generating meta-analytic estimates of mean relative facet contributions and their variability across units of contextual factors are presented in Appendix C.

Upon generating the meta-analytic estimates discussed above, one may use a variety of common methods to determine whether the "true" relative contribution of a given facet substantially varies across units of the contextual factor. One method would be to calculate the $Q$ statistic (Hunter & Schmidt, 1990), which is:

$$Q = K \left[ \frac{\text{Var(\text{raw})}}{\text{Var(\text{error})}} \right]$$

(39)

In this equation, $Q$ is distributed as $\chi^2$ with $K-1$ degrees of freedom, where $K$ is the number of units of the contextual factor being examined (e.g., number of organizations). $\text{Var(\text{raw})}$ refers to the raw variance of relative facet contributions across units the contextual factor. Lastly, $\text{Var(\text{error})}$ refers to the amount of the aforementioned raw variance that can be attributed to sampling error. One $Q$ test can be conducted for each facet-by-contextual factor combination,
testing the null hypothesis that all VAA facet contributions are equal across contextual factor units.

Alternatively, another way one may assess whether stability likely exists in facet contributions across units of the contextual factor is to rely on Hunter and Schmidt's (1990) "75% rule." Although this "rule of thumb" is applied in the context of validity generalization in personnel selection (e.g., Guion, 1998), it appears reasonable to apply in this case as well because the concern is basically whether facet contributions of a construct's variance architecture generalize from one organizational, cultural or temporal condition to another. Hunter and Schmidt's (1990)'s 75% rule states that if known sources of artifactual variation (e.g., sampling error) account for 75% or more of the variance in effect size (in this case, relative facet contribution) estimates across studies (in this case, units of the context factor), than the situational specificity hypothesis may be rejected. In terms of the VAA, rejection of such a hypothesis would suggest that true differences do not exist in a facet's contribution across the units of the given context factor.

Comparing Variance Architectures

The information stemming from constructs' variance architectures is quite rich, and provides the opportunity to make several types of comparisons between variance architectures underlying constructs, their measures, and dimensions (in the case of multidimensional constructs). As introduced earlier, the primary information that results from delineating a construct's variance architecture is the absolute and relative amounts of total construct variance accounted for by each VAA facet and the relative stability of those contributions across contexts and time. In light of this resulting information, three primary dimensions along which variance architectures can be compared emerge. Figure 2 graphically depicts these dimensions.
The three dimensions along which variance architectures can be compared are (a) form of comparison, (b) level of comparison, and (c) unit of comparison. The “form” of comparison dimension differentiates comparisons of facet contributions and their stability across contexts based on absolute and relative differences. For example, one can compare architectures for differences in the variance components stemming from the facets they have in common (i.e., absolute facet contributions to total variance) or one can compare architectures for differences in the relative contribution of their common facets to total variance. The form of comparison dimension is represented by the axis that is anchored by absolute and relative comparisons in Figure 2.

![Figure 2. A Three-dimensional Framework for Comparing Variance Architectures](image)

The “level” of comparison dimension differentiates comparisons of facet contributions and their stability across contexts made at the architecture- and facet-levels. For example, one can
make holistic, architecture-level comparisons of profiles of absolute or relative VAA facet contributions to total variance. Alternatively, one can make specific facet-level comparisons, examining facet contributions or their stability in a pair-wise fashion, focusing on differences between single facets one at a time. The level of comparison dimension is represented by the axis that is anchored by architecture- and facet-level comparisons in Figure 2.

Last, the “unit” of comparison dimension differentiates between comparison of facet contributions and facet stabilities across contexts. For example, when examining two architectures one can compare facet contributions (both absolute and relative) and the stability of those contributions across contexts. The unit of comparison dimension is represented by the axis that is anchored by facet contribution and facet stability in Figure 2.

Crossing these three dimensions along which architectures can be compared provides eight basic types of comparisons that can be made of two or more constructs’ variance architectures. In the sections that follow, I systematically review (a) the eight basic types of comparisons, (b) conditions under which these comparisons are appropriate, and (c) potential methodology for making the comparisons. The following sections are organized by level of comparison because it is the dimension across which the methods for comparing variance architecture information are most divergent.

Architecture-Level Comparisons

Four types of architecture-level comparisons are possible within the framework revealed by Figure 2. I will refer to these comparisons as Types I-A through IV-A, where the “A” signifies that they are comparisons made at the architecture-level. Type I-A and II-A both deal with architecture-level comparisons of patterns of facet contributions to total construct variance. Specifically, Type I-A comparisons focus on comparing profiles of absolute facet contributions to total construct variance (i.e., comparing profiles of variance component values), whereas Type II-
A comparisons focus on comparing profiles of *relative* facet contributions to total construct variance (i.e., comparing profiles of proportions of total construct variance accounted for by each facet). Type III-A and IV-A both deal with architecture-level comparisons of patterns of facet stabilities. Specifically, Type III-A comparisons focus on comparing profiles of "true variance" around absolute facet contributions across contexts, whereas Type IV-A comparisons focus on comparing profiles of "true variance" around relative facet contributions across contexts.

Making architecture-level comparisons is only appropriate if the variance architectures being compared are based on the same architectural design. Specifically, the two variance architectures being compared should be composed of the same facets and those facets should be oriented to each other in the same manner in both architectures (i.e., both architectures have the same pattern of nesting and crossing among their facets). As illustrated in earlier sections, including a facet in one architecture and leaving it out of another can substantially alter the apparent amount of variance accounted for by other facets in an architectural design (see Example 2, page 64). Thus, when comparing two architectures, it is important that the architectures be of the same design unless one has strong theoretical reasons why differences between designs would not alter estimates of the contribution or stability of facets that the designs share in common.

Furthermore, when making comparisons of Type I-A and III-A (those dealing with comparisons of absolute facet contributions and their stabilities) the architectures being compared should be based on constructs that were assessed on the same metric (e.g., a 1 to 7 scale). For example, it would make little sense to compare the profiles of absolute facet contributions of one construct that was assessed on a 100-point scale to one that was assessed on a 5-point scale (i.e., a Type I-A comparison). The facet contributions based on the 100-point scale would almost necessarily be higher due to differences in scaling. For this latter reason, making comparisons of
Type I-A or III-A between variance architectures may be limited, but is still be possible if constructs, or more likely, dimensions of the same construct are assessed using a similar metric.

There are two aspects of Type II-A and IV-A comparisons (those dealing with comparisons of relative facet contributions and their stabilities) that make them particularly attractive relative to Type I-A and Type IV-A comparisons. First, constructs do not have to be assessed on the same metric for their variance architectures to be compared. This is because the values that are being compared across constructs are proportions of total construct variance, and within each architecture those components will sum to one, thus essentially putting the values to be compared on the same metric. Second, a metric can be established for making Type II-A and IV-A comparisons that indexes the similarity profiles of relative facet contributions and stability that ranges from 0 to 1 no matter how the underlying constructs of each variance architecture are scaled. The fact that such a scale invariant index can be created for evaluating the similarity of variance architectures provides researchers with a tool for comparing results of VAA across many different studies. For these reasons, at the architecture-level, I focus primarily on describing methods of making comparisons of Type II-A and IV-A.

Unfortunately, making architecture-level comparisons of any type are difficult due to the lack of available inferential methods for making such comparisons in the current statistical literature. Indeed, inferential statistical techniques are not even that well developed for comparing the contribution of single facets of constructs' variance architectures (either absolute or relative) in a pairwise manner, let alone sets of such values that constitute constructs' variance architectures (McGraw & Wong, 1996; Schroeder & Hakstian, 1990). Thus, until such methods are further developed, the best strategy one has for holistically comparing such information is to compute simple descriptive indexes of the similarity of variance architectures, is detailed below.
Making comparisons of Type II-A can be achieved through the formation of a simple
descriptive index that reflects the similarity of two architectures in terms of their profiles of
relative facet contributions. This index, which I refer to as the architecture similarity index for
relative facet contributions (\textit{ASI-RC}) can be formed by taking a function of the mean of absolute
differences between the relative contributions of their corresponding facets. The ASI-RC index is
calculated via the following formula:

\[
ASI-RC = \sqrt{\frac{\sum_{x=1}^{m} p_{\sigma_{x}^{2}} - \frac{2}{m}}{m}} - \frac{2}{m}
\]

In this formula, \(p_{\sigma_{x}^{2}}\) represents the proportion of variance accounted for by the \(x^{th}\) facet
or interaction term of the first (\(j = 1\)) or second (\(j= 2\)) architecture considered. The \(m\) in this
formula refers to the total number of facets and interactions being estimated in the architectural
design under consideration.\(^{19}\) Regardless of how the constructs underlying each architecture are
scaled, such an index would have an upper bound of 1, indicating the relative contribution of all
corresponding facets for the two architectures are identical, and a lower bound of 0, indicating the
maximum possible difference between the two architectures in terms of profiles of their facets
relative contributions was achieved.

Although the above index addresses architecture-level comparisons of Type II-A, it does
not provide an indication of the similarity of architectures in terms of their stability of their
facets’ relative contributions to construct variance (i.e., Type IV-A comparisons). To make such
comparisons one can form a descriptive index that is a function of the mean of absolute
differences between the stability of the relative contribution of corresponding facets of two
architectures across the units of contextual factor of interest (e.g., organizations, cultures, or times). I will refer to this index as the architecture similarity index for the stability of relative facet contributions (ASI-RS), and calculate it via the following formula:

\[
ASI-RS = 1 - \left( \sum_{x=1}^{m} ISDR_{p_{x}}^{2} \cdot \left( 1 - \frac{\text{Var}(z_{\sigma}^{2})_{\text{error}_{-x1}}}{\text{Var}(z_{\sigma}^{2})_{\text{raw}_{-x1}}} \right) \right) - \left( \frac{m}{m} \right) ISDR_{p_{x}}^{2} \cdot \left( 1 - \frac{\text{Var}(z_{\sigma}^{2})_{\text{error}_{-x2}}}{\text{Var}(z_{\sigma}^{2})_{\text{raw}_{-x2}}} \right) \]

(41)

As with the formula for ASI-RC, the \( m \) in this formula refers to the total number of facets and interactions being estimated in the architectural design under consideration. Further, “\( ISDR_{p_{x}} \)” represents the inter-standard deviation range of the proportion of total variance accounted for by the \( x \)th facet or interaction term of the \( j \)th architecture across the units of the contextual factor under consideration. Thus, the \( ISDR_{p_{x}}^{2} \) value is similar to the inter-quartile range of a facet’s relative contribution across units of the contextual factor, except that its bounds are plus and minus one standard deviation instead of the first and third quartiles. Generating raw standard deviation estimates for relative facet contributions across units of the contextual factor in question are based on the meta-analytic described in Appendix C. The quantity by which the ISDR value is multiplied reflects the square root of the proportion of raw relative facet contribution variance across units of the contextual factor that reflects true variance in relative facet contributions across contexts. This quantity can also be obtained from using meta-analytic methods detailed in Appendix C. Thus, multiplying the ISDR value times this quantity for any given facet in any given architecture gives one an index of the true amount of variation (after removing artifactual variation due to sampling error) for a given facet across units of the contextual factor. These values are computed for each facet of the two architectures being
compared and the absolute value of the differences between corresponding facets are averaged across all facets in the designs to obtain the ASI-RS index.

Like the ASI-RC index, the ASI-RS index is scaled such that higher numbers indicate architectures are more similar. The index essentially reflects one minus the mean absolute difference between indexes of the true variability in VAA facets across units of the contextual factor in question (expressed in raw, not squared units). The fact that ISDR values may only range from 0 and 1, puts bounds of 0 and 1 on the ASI-RS index. As such, the ASI-RS index’s upper bound of 1 indicates that the stability of all corresponding pairs of architectural facets for the two constructs is identical, and its lower bound of 0 indicates that the stability of all corresponding pairs of architectural facets for the two constructs is maximally different.

Facet-Level Comparisons

Unlike architecture-level comparisons, the foci of facet-level comparisons are not profiles of facet contributions or their stability, but rather single facet contributions and their stability. Four types of facet-level comparisons are possible within the framework revealed by Figure 2. I will refer to these comparisons as Types I-F through IV-F, where the “F” signifies that they are comparisons made at the facet-level. Type I-F and II-F both deal with the comparison of single facet contributions to total construct variance. Specifically, Type I-F comparisons focus on comparing the absolute contribution of a single facet to total construct variance (i.e., comparing variance component values of a single facet in two different architectures), whereas Type II-F comparisons focus on comparing the relative contribution of a single facet to total construct variance (i.e., comparing proportions of total variance accounted for by a single facet in two different architectures). Type III-F and IV-F both deal with comparisons of the stability of a single facet contribution to total construct variance. Specifically, Type III-F comparisons focus on comparing the "true variance" around the absolute contribution of a single facet across contexts,
whereas Type IV-F comparisons focus on comparing the "true variance" around the relative contribution of a single facet across contexts.

As was the case with making architecture-level comparisons, making facet-level comparisons is only appropriate if the variance architectures being compared are based on the same architectural design (for reasons stated earlier). Similarly, when making comparisons of Type I-F and III-F (those dealing with comparisons of absolute facet contributions and their stability) the architectures being compared should be based on constructs that were assessed on the same metric. Thus, as was the case with architecture-level comparisons of Types I-A and III-A, making facet-level comparisons of Types I-F and III-F may be limited, but may still be possible if constructs, or more likely, dimensions of the same construct are assessed using a similar metric.

Although methods for making inferential architecture-level comparisons between constructs' variance architectures are scant, some methods do hold promise for making inferential comparisons between single facets of variance architectures. For example, one can make inferential comparisons of Type I-F by forming confidence intervals around the variance component for a given facet in one design (i.e., the facet’s absolute contribution) and comparing it to the variance component associated with that facet in the other design. Similarly, one can make inferential comparisons of Type II-F by forming confidence intervals around the proportion of variance accounted for by a given facet in one design (i.e. the facet’s relative contribution) and comparing it to the proportion associated with that facet in the other design. Confidence intervals surrounding a facet’s relative contribution to total construct variance can be constructed using standard errors based on the normalizing transformation alluded to earlier and described in Appendix B. Because these both absolute and relative facet contributions are not normally distributed, the confidence intervals surrounding them will tend to be asymmetrical.
Comparing the stability of single facet contributions (both absolute and relative) is a bit trickier than comparing the magnitudes of the facet contributions themselves. As alluded to earlier, it is not even possible to make Type III-F comparisons because there is no method currently available for estimating true variance of absolute facet contributions across units of one’s contextual factor of interest. One method for indirectly making Type IV-F comparisons would be to simply determine whether that facet was deemed stable by the methods described earlier for each construct separately (e.g., Hunter and Schmidt's Q test, or 75% rule) and then compare results. A more direct inferential test would be to take the true variances estimates for a facet’s relative contribution across contexts (derived metaanalytically) and use an F-test to test such facet variances from each architecture for equality (e.g., Guilford & Fruchter, 1973).

Summary of Variance Architecture Comparisons

As the above discussion revealed, one may make eight basic types of comparisons among variance architectures that can be described along three dimensions. Given the numerous types of comparisons introduced above, I have summarized descriptions of the eight basic types in Table 4. Although eight comparisons were introduced, two of these (i.e., III-A and III-F) are not currently feasible due to a lack of methodology for estimating the true variance in absolute facet contributions across units of one’s’ contextual factor of interest.

In introducing architecture-level comparisons the emphasis was on relative comparisons due to the availability of the fact that constructs being assessed did not need to be on the same scale, as well as availability of an index with a metric that was scale invariant, both of which were cited as potential drawbacks of making absolute architecture-level comparisons. At the facet-level no such judgment was offered. Depending on the question one wishes to answer, one may wish to compare either absolute- or relative-facet contributions using inferential methods for making both forms of comparison.
Table 4. A Summary of the Eight Basic Types of Variance Architecture Comparison

<table>
<thead>
<tr>
<th>Type</th>
<th>Level</th>
<th>Form</th>
<th>Unit</th>
<th>Focus of Comparison</th>
<th>Similarity Index</th>
<th>Conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>I-A</td>
<td>Architecture</td>
<td>Absolute</td>
<td>Contribution</td>
<td>Profile of absolute VAA facet contributions to construct variance</td>
<td>-</td>
<td>constructs on same metric, same architectural design</td>
</tr>
<tr>
<td>II-A</td>
<td>Architecture</td>
<td>Relative</td>
<td>Contribution</td>
<td>Profile of relative VAA facet contributions to construct variance</td>
<td>ASI-RC</td>
<td>same architectural design</td>
</tr>
<tr>
<td>III-A*</td>
<td>Architecture</td>
<td>Absolute</td>
<td>Stability</td>
<td>Profile of &quot;true variance&quot; around absolute VAA facet contributions across contexts</td>
<td>-</td>
<td>constructs on same metric, same architectural design</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Profile of &quot;true variance&quot; around relative VAA facet contributions across contexts</td>
<td>ASI-RS</td>
<td>same architectural design</td>
</tr>
<tr>
<td>IV-A</td>
<td>Architecture</td>
<td>Relative</td>
<td>Stability</td>
<td>&quot;True variance&quot; of a single facet's relative contribution across contexts</td>
<td>-</td>
<td>same architectural design</td>
</tr>
<tr>
<td>I-F</td>
<td>Facet</td>
<td>Absolute</td>
<td>Contribution</td>
<td>Absolute contribution of a single facet to construct variance</td>
<td>-</td>
<td>constructs on same metric, same architectural design</td>
</tr>
<tr>
<td>II-F</td>
<td>Facet</td>
<td>Relative</td>
<td>Contribution</td>
<td>Relative contribution of a single facet to construct variance</td>
<td>-</td>
<td>same architectural design</td>
</tr>
<tr>
<td>III-F*</td>
<td>Facet</td>
<td>Absolute</td>
<td>Stability</td>
<td>&quot;True variance&quot; of a single facet's absolute contribution across contexts</td>
<td>-</td>
<td>constructs on same metric, same architectural design</td>
</tr>
<tr>
<td>IV-F</td>
<td>Facet</td>
<td>Relative</td>
<td>Stability</td>
<td>&quot;True variance&quot; of a single facet's relative contribution across contexts</td>
<td>-</td>
<td>same architectural design</td>
</tr>
</tbody>
</table>

*No method currently exists for generating true variance estimates of a facet’s absolute contribution to construct variance across contexts. As such, these two basic forms of architecture comparison are not yet feasible, but are recognized because they naturally fit into the framework introduced in this study. “True variance” statements are delimited by quotation marks to signify that these variances are estimates of true variance across contexts derived from meta-analytic methods (see Appendix C).
The Variance Architecture Approach: Summary

The variance architecture approach is a research methodology that offers much promise to organizational researchers. In this section, a review of the VAA is provided for purposes of general summary.

The historical roots of the VAA lies in both Cattell's (1966) Basic Data Relations Matrix as well as Cronbach et al.’s Generalizability Theory (1963). Borrowing from the perspectives on variance these two frameworks provide and the literatures of I/O psychology, organizational behavior, cross-cultural psychology, and psychometrics, the VAA was formulated. The core ideas of the variance architecture approach center on the notion that (a) all constructs that arise in organizational contexts can be described in terms of their patterns of variation across a common set of facets along which they all may vary, and (b) ascertaining information regarding the patterns of variance and their relative stability could provide a number of benefits for organizational science and practice. Nine facets of variation were introduced as the basis of any construct's variance architecture, namely culture, organization, group, job, individual, observer, item, stimulus, and occasion facets.

Once this content-related foundation of the VAA was established, six primary benefits of the VAA for organizational science and practice were introduced. Specifically, I discussed how the VAA can (a) identify areas where future theory-building, prediction, and intervention efforts might have most / least impact for a given construct, (b) estimate likely upper bounds of variance accounted for by predictors stemming from a certain level for a given construct, (c) clarify appropriate levels of analysis for a given construct, (d) identify the degree of stability in the dispersion of variance across facets in a given construct, (e) provide a unified framework for comparing different: constructs, methods of measuring a single construct, and dimensions of a multidimensional construct across multiple levels of analysis, and (f) indicate where areas of the
research literature are scant with regard to information regarding potential facets of construct variation.

Upon describing the aforementioned benefits of the VAA, the next issue addressed was methods for delineating a construct’s variance architecture. The VAA research process was characterized as a three-step process consisting of (a) data gathering, (b) specifying architectural designs based on the data currently available, and (c) estimating parameters for the architectural designs specified using the GLMM and examining their stability using methods derived from the meta-analysis literature. Finally, a unified framework for comparing variance architectures was introduced that delineated eight types of comparisons that could be made. Figure 3 provides a graphic summary of key aspects of the VAA as detailed in the sections presented above.

Figure 3. A Summary of the Variance Architecture Approach to the Study of Constructs
Although exposition of the VAA is the primary purpose of this dissertation, to help elucidate many of the issues discussed above, it will be useful to demonstrate how the VAA research process unfolds with an example from the applied literature. By using raw data from the current Organizational Citizenship Behavior (OCB) literature, I provide a concrete examples of the type of (a) information that would result from the VAA approach, (b) answers that could be provided for research questions related to this approach, and (c) problems and benefits that may arise working through various aspects of the VAA research process.

Organizational Citizenship Behavior through a VAA Lens

Katz and Kahn (1978) proposed that three classes of employee behavior are necessary for the effective functioning of an organization. The first class of behavior regards employees fulfilling the duties proscribed by their formal job descriptions. The second class of behavior regards employees staying with the organization. The third class of behavior regards employees engaging in “innovative and spontaneous behavior: performance beyond role requirements” (p.332). Without this latter class of cooperative acts the social fabric of organizations and the groups that compose them would breakdown (Katz & Kahn, 1978). Within the I/O and organizational behavior literature, this latter class of behaviors has come to be studied under the rubric of Organizational Citizenship Behaviors (OCB; Bateman and Organ, 1983; Organ, 1988; Podsakoff, MacKenzie, Paine, & Bachrach, 2000).

Definitions and Examples of OCB

Organ (1988) formally defined OCB as "individual behavior that is discretionary, not directly or explicitly recognized by the formal reward system, and that in the aggregate promotes the effective functioning of the organization” (p. 4). Organ (1988) goes on to clarify some aspects of this definition by adding:
"By discretionary we mean that the behavior is not an enforceable requirement of the role or the job description, that is, the clearly specifiable terms of the person's employment contract with the organization; the behavior is rather a matter of personal choice, such that its omission is not generally understood as punishable" (p.4).

Examples of behavior that reflect OCB are: "helping fellow co-workers who have heavy workloads," "attending meetings that are not required, but help the company image," and "trying to avoid creating problems for coworkers" (Organ, 1988).

Typically, OCB has been viewed as a multidimensional construct consisting of five dimensions, namely altruism, civic virtue, conscientiousness, courtesy, and sportsmanship (e.g., Podsakoff, MacKenzie, Moorman, & Fetter, 1990). Altruism reflects OCB directed towards helping or assisting one's co-workers with their work-related problems, duties or questions (e.g., helping others who have been absent). Civic Virtue refers to behavior reflecting active involvement, participation, and concern for the life of the organization (e.g., providing constructive suggestions to improve the organization, keeping up with organizational announcements). Conscientiousness reflects behavior that goes above and beyond the minimum requirements set forth by organizational rules, policies, and procedures (e.g., having attendance at work that is above the norm, not taking extra breaks). Courtesy refers to behaviors aimed at preventing work-related problems from occurring with other employees (e.g., considering the impact of one's actions on his or her co-workers, not abusing the rights of others). Lastly, Sportsmanship refers to behaviors that reflect a person's tendency to tolerate less than ideal organizational circumstances that may arise in the course of one's work without complaining (e.g., tending to make "mountains out of molehills," consuming a lot of time complaining about trivial matters).
Although the five-factor conceptualization of OCB is most common, researchers have also conceptualized the OCB construct in terms of two general factors that subsume the five factors described above (Organ, 1994; Williams & Anderson, 1991). Namely, researchers have distinguished between OCB directed at the organization or primarily beneficial for the organization (OCBO), and OCB directed at specific individuals within the organization or with primary benefit for specific individuals (OCBI). Due to their focus on the helping and avoidance of problems with other workers, Altruism and Courtesy OCB are subsumed under the OCBI factor. Conversely, Civic Virtue, Conscientiousness, and Sportsmanship, all deal with behaviors that benefit the organization in general and, thus, are considered representative of the OCBO factor (Williams & Anderson, 1991).

The Importance of Examining OCB

The importance of studying OCB is made apparent by its interest to researchers in several different areas of the I/O / OB literature. Three primary areas of research with regard to OCB have formed in the literature, which I briefly review below.

One area of research with regard to OCB has focused on investigating the implications of employees' OCB engagement for organizational effectiveness (e.g., Podsakoff et al., 1997; Podsakoff and MacKenzie, 1994; Walz & Niehoff, 1996). This body of literature has identified several reasons how OCB may benefit organizations, such as:

"(a) enhancing coworker and managerial productivity, (b) freeing up resources so they can be used for more productive purposes, (c) reducing the need to devote scarce resources to purely maintenance functions, (d) helping to coordinate activities both within and across workgroups, (e) strengthening the organization's ability to attract and retain the best employees, (f) increasing the stability of the
organization's performance, and (g) enabling the organization to adapt more effectively to environmental changes." (Podsakoff et al., 2000, p. 545).

Although empirical evidence emerging from this area of research suggests the impact of OCB on organizational effectiveness (e.g., indexes of group level productivity like total sales, quality indexes, customer satisfaction) is quite positive, it appears that some forms of OCB (e.g., Altruism), may be more beneficial than others (e.g., Sportmanship and Civic Virtue; Podsakoff et al., 2000).

Another area of research with direct ties to OCB has focused on mapping the job performance construct domain (e.g., Borman & Motowidlo, 1993; Campbell, McCloy, Oppler, & Sager, 1993; Conway, 1996). This research has consistently revealed that job performance consists not only of traditional job-specific, on-task performance behaviors, but also, nonspecific contextual performance behaviors such as OCB (e.g., Campbell, 1990; Motowidlo and Van Scotter, 1994). A related body of research has investigated the implication that broadening the job performance construct to include elements of OCB has for personnel functions, namely personnel selection and performance assessment (e.g., Borman & Motowidlo, 1997; Hattrup, Rock, & Scalia, 1997; Murphy & Shiarella, 1997). Such research has generally demonstrated that the validity of a given selection test may vary, depending on how much weight is given to OCB-type behaviors, as opposed to traditional performance behaviors when formulating a performance criterion (e.g., Murphy & Shiarella, 1997). Specifically, findings indicate that personality variables increase in their criterion-related validity when aspects of performance that tap OCB are given more weight, which has implications not only for validity, but adverse impact issues as well (e.g. Hattrup et al., 1997; McHenry, Hough, Toquam, & Hanson, 1990).

Lastly, the largest area of OCB research regards the study of antecedents of employees' engagement in such behavior (e.g., Konovsvky & Organ 1996; Morrison, 1994; Moorman, 1991;
Williams & Anderson, 1990). Given that OCB has traditionally been viewed as an individual-level construct, most of the empirical research that has examined antecedents of employees' engagement in OCB has focused on individual-level predictors of it (Organ & Ryan, 1995; Podsakoff et al., 2000). However, employees' engagement in OCB has also been found to significantly vary at other levels, and has been tied to specific predictors at those levels. For example, multiple variables have been found to be predictive of OCB at the workgroup-level (e.g., Pearce & Gregersen, 1991; George, 1990; George & Bettenhausen, 1990; Kidwell et al., 1997). Additionally, significant variation in OCB engagement has also been found at department- and organization-levels, and predicted by variables stemming from those levels (Pearce & Gregersen, 1991; Schnake & Dumler, 1997; Randall, Cropanzano, Borman, Birjulin, 1999; Wright, George, Farnsworth, & McMahan, 1993). Still others have found evidence of substantial item- or behavior-related variation and antecedents (e.g., Morrison, 1994; Pond, Nacoste, Mohr, & Rodriguez, 1997; Putka & Vancouver, 1999), as well as culture-level variation (e.g., Fahr, Earley, & Lin, 1997; Lam et al., 1999). Although research has taken place within these various levels, and significant predictors of OCB have been found that are isomorphic with these levels, what is not clear is the relative importance that these significant findings may have in the larger scheme of understanding employees' engagement in OCB when one considers all levels simultaneously. Thus, by examining OCB engagement through a VAA research lens, I will help to provide a context for understanding the relative importance of the specific empirical findings that have emerged at several levels of analysis in the OCB literature.

**OCB-Related Research Questions**

Earlier in this introduction of the VAA, several benefits of the VAA were described and examples of research questions that could be addressed based on VA-related information were presented in Table 2. In the present investigation of OCB through the VAA lens, many of the
questions listed in Table 2 were answered as they pertain to the OCB construct. In addition to these general questions, several questions, specific to the OCB construct were addressed by information resulting from the VAA. The OCB-specific questions examined as part of the present investigation are reviewed below.

The Adequacy of VAA-related Data in the OCB Literature

Before any empirical OCB data was analyzed, I addressed three diagnostic questions that assessed whether the OCB literature provided the data necessary to provide a full delineation of the OCB construct’s variance architecture. First, does the OCB research literature contain data that are sampled across all VAA facets? For example, the literature on OCB contains data from hundreds of organizations, across a wide array of industries (e.g., Podsakoff et al., 2000), however, it contains far fewer studies that have examined OCB in other cultures (e.g., Lam et al., 1999), or studies that examine the same individuals on multiple occasions (e.g., Bateman & Organ, 1983). Examples such as this suggest that the OCB literature may lack sufficient data to provide a description of OCB’s variance architecture that includes a culture facet.

The second diagnostic question I addressed was as follows: Does the OCB research literature contain data that are coded in such a way that they allow for identification of the units of the VAA facets from which they are sampled? For example, even though studies may have examined several individuals’ engagement in OCB dispersed across a wide variety of groups in a particular organization, it may be the case that researchers failed to record information regarding which individual stemmed from which workgroups (e.g., more likely if the individual-level of analysis was the focus). In these cases, information on the group-facet would not be coded into the data, and one would be unable to distinguish one group from another within the data set. To the extent that this is the case, it would hamper one's ability to provide a detailed description of OCB's variance architecture.
The third diagnostic question I addressed was as follows: Is the OCB research literature characterized by measurement and sampling designs that allow for the unique estimation of each VAA facet’s contribution to variance in the OCB construct? As discussed earlier, many organizations are hierarchically nested entities (e.g., individuals within jobs within groups within departments, etc.). Such nesting limits one's ability to uniquely estimate the contribution of individual facets. Thus, without variation from this “nested norm,” it may be difficult to provide unique estimates for the contribution of each VAA facet and its interaction with other facets to the variance in the OCB construct. Examining the extent to which confounding of facet effects occurs in the architectural designs that are made possible by the data available in the OCB literature was the focus of this last diagnostic question.

Providing answers to these initial diagnostic questions was achieved by gathering data from the currently existing OCB literature and laying out as detailed an architectural design on the variance in OCB that the data currently allows. In the process of attempting to lay out designs based on the OCB data available, answers to the three diagnostic questions posed above became readily apparent (detailed in the methods section). The benefit that answering such questions is that they can inform OCB researchers where gaps in our knowledge exist regarding the variance in the OCB construct.

Architectures of the OCB Construct in General

Due to the uncertainty regarding the data that the OCB literature would be able to provide for the current investigation of the variance underlying the OCB construct, the quality with which all research questions introduced below depended on the data available to answer them. Thus, answers to subsequent questions were limited based on the data available in the empirical OCB literature, and how much of it was acquired for this study.
The general questions that I attempted to address in this investigation of the OCB construct stem directly from the discussion of VAA benefits addressed earlier. Specifically, this investigation attempted to determine (a) where the most/least variance in the OCB construct lies, (b) the upper bounds on the predictive ability of specific variables stemming from a particular facet of OCB’s variance architecture, and (c) how stable facet contributions to the variance in the OCB construct are across different organizations, cultures, and times.

Unfortunately, because little longitudinal or cross-cultural OCB research data existed at the time of this investigation, it was not feasible to evaluate the contribution of the culture or occasion facets to the variance in OCB. Moreover, such a lack of data also made investigating the stability of facet contributions across cultural contexts and time difficult. With regard to the stability of facet contributions across time periods, the OCB literature only formally came into existence in 1983, with the publication of two seminal empirical articles (e.g., Bateman & Organ, 1983; Smith, Organ, & Near, 1983), thus, examining stability of across time periods (e.g., one year or five year blocks) was limited by the degree to which data gathered is dispersed across the past few decades. Although examining the stability of architectures across cultures or time periods was problematic in the present investigation, examining the stability of architectures across organizations was possible due to the large number of organizations across which OCB has been examined (Podsakoff et al., 2000).

Architectures of OCB Dimensions

As cited earlier, one of the benefits of the VAA is that it provides a unified framework for comparing not only different constructs, but also different measures of the same construct, as well as different dimensions of a multidimensional construct. In the present investigation of OCB, this benefit of the VAA can be nicely illustrated because is OCB a multidimensional construct and is often assessed by a variety of methods. Because of these attributes, several sets of questions
regarding how architectures of OCB dimensions compare to one another, as well as how architectures revealed by different methods for measuring OCB compare were addressed.

Based on the OCB literature, there appears to be little justification for making a-priori hypotheses regarding how variance architectures underlying the various dimensions of OCB may differ. Indeed, in their recent review of the OCB literature, Podsakoff and colleagues (2000) reflect such uncertainty by suggesting that more research is need with regard to the investigation of similarities and differences between different dimensions of OCB. In light of this need, this investigation explored the possibility that dimensions of OCB differ in their variance architectures. This was achieved by examining the architectures underlying each dimension of OCB separately.

In addition to comparing the variance architectures underlying the dimensions of the OCB construct for areas of similarity and difference, I addressed two research questions. First, does accounting for the multidimensionality of the OCB construct reduce the amount of variation attributable to behavior-related facets in architectures that ignore such multidimensionality? If accounting for the multidimensionality of does not reduce the amount of variance in OCB that is attributable to behavior-related facets (e.g., behavior-facet, individual-by-behavior interaction), than it is not likely important to fit separate architectures for each dimension (Shavelson & Webb, 1991). A primary reason to expect a drop in variance among the behavior-related facets and, in particular, the individual-by-behavior interaction, is that different individual-level predictors have been found to differentially predict different dimensions of OCB (e.g., Organ & Ryan, 1995; Podsakoff et al., 2000).

Answering this question requires the comparison of two architectures that are based on OCB data across all dimensions. First, I examined the variance architecture underlying OCB in general (across all behaviors), temporally ignoring the dimensionality of the OCB construct.
Upon delineating this general architecture, a subsequent architecture was delineated that includes OCB dimension as a random source of variation in the model. Adding in dimension as a factor allows one to model the impact of the individual-by-dimension interaction, which may account for a large portion of any individual-by-behavior variance (cited above) found in the initial model where dimension is ignored. Specifically, if comparison of the behavior-related facets stemming from both architectures revealed that the contribution of the behavioral-related facets from the initial model (e.g., the behavior facet, and behavior-by-other facet interaction terms) are significantly reduced when dimension is included in the model, and the model with dimension provides a significantly better fit to the data, then this is an indicator that decomposing the variance in OCB for each dimension separately is justified (Greguras & Robie, 1998; Shavelson & Webb, 1991). Such a finding would offer support for claims of the multidimensionality of the OCB construct.

The second dimension related question that was addressed in the present investigation is as follows: Are the architectures of the OCB dimensions nested within each higher-order OCB dimension (i.e., OCBI and OCBO) more similar to each other than the architectures of the OCB dimensions in the other higher-order OCB dimension. For example, are variance architectures underlying Altruism and Courtesy (subsumed within the higher-order OCB dimension, OCBI) more similar to each other than the variance architectures underlying Civic Virtue, Conscientiousness, and Sportsmanship (subsumed within the higher-order OCB dimension, OCBO)? Such a finding would lend further support to the two-factor model of OCB engagement, as well as provide information with regard to the level at which the most differentiation between OCBI and OCBO (in terms of variance accounted for by a given facet) may occur. To answer this question, I estimated the similarity of architectures underlying each dimension of OCB by comparing their corresponding ASI-RC and ASI-RS values (introduced earlier).
Architecture of OCB Revealed by Different Measurement Methods

In addition to addressing differences between the architectures underlying dimensions of OCB, the present investigation also explored how architectures of OCB revealed by different methods of assessing the OCB construct compare to one another. Individuals' engagement in OCB has usually been assessed in three ways. OCBs are most often assessed by having supervisors rate each of their subordinate employees on each of several behaviors that are representative of the OCB domain. Less often, peer and self-ratings of individuals' engagement in OCB are provided in a similar manner. Based on the relative strengths and weaknesses of these rating sources revealed by decades of research in the performance appraisal literature (e.g., Murphy & Cleveland, 1995), there are several reasons to believe that these rating sources might differ in terms of the variance architectures of the OCB construct they reveal. Three sets of questions were examined regarding differences between OCB architectures revealed by these three methods of measurement. Specifically, questions regarding: (a) self-peer differences, (b) self-supervisor differences, and (c) peer-supervisor differences were addressed.

Self-Peer Differences. Based on the performance appraisal and social psychology literatures, there are several differences that one might expect to observe between the architectures revealed by self and peer ratings of OCB. For example, one might expect that the individual, job, group, and organization facets to contribute less variance among self-ratings compared to peer ratings. The reason for this is that self-ratings have commonly been criticized in the performance appraisal literature for being inflated estimates of an individual's true level of performance, stemming from individuals' tendency to exhibit a self-enhancing bias (Murphy & Cleveland, 1995). Such a bias would likely lower the ceiling on the potential contribution of variance by these facets in self-ratings relative to peer ratings.
A second difference that might be expected between architectures revealed by self and peer ratings is that the individual-by-behavior interaction and behavioral sources of variation in general, may be greater for self-ratings relative to peer ratings. There are several reasons to expect such differences, which may be exacerbated depending on how peer ratings of OCB engagement were gathered. One reason to expect a difference is that self-raters are likely to have less of a cognitive load in terms of the rating task relative to their peers, particularly if their peers are rating multiple individuals (Lord & Maher, 1991). With less of a cognitive load, self-raters may be better able to make finer distinctions between specific instances of OCB, which may lead to more behavior-related variation among self-raters.

Another reason behavior-related facets may contribute more to variance in self-ratings relative to peer ratings is that self-raters may have a greater motive to distinguish between specific instances of OCB relative to their peers. Specifically, self-raters may have a greater tendency to inflate their ratings of engagement in OCB on those specific behaviors that make them appear most effective (e.g., Jones & Nisbett, 1971; Mitchell, et al., 1981). The impact such a tendency might have on architectures of OCB revealed by self-ratings is that (a) a greater contribution of the individual-by-behavior interaction may be observed to the extent that individuals disagree about the extent to which various OCB make them appear effective, or (b) a greater contribution of the behavior facet may be observed to the extent that individuals generally agree about the extent to which various OCB make them appear effective. Although the present study did not examine whether individuals’ actually agree with regard to the degree to which various OCBs make them appear effective, past studies have indicated that individuals’ perceptions of OCBs with regard to whether they are necessary of effective performance on the job suggest that very little agreement exists among employees (e.g., Putka & Vancouver, 1999). Based on the arguments presented above, such findings suggest that the impact of individual-by-
behavior interaction may contribute more variance to self-ratings of OCB than the behavior facet.

*Self-Supervisor Differences.* Differences between architectures revealed by self and supervisor ratings are likely to be very similar to those between self and peer ratings, however, differences in architectures between self and supervisor ratings are likely to be even more severe. For example, one might expect the individual facet to account for more variance among supervisor ratings relative to self-ratings and behavior-related sources of variation (e.g., behavior facet, individual-by-behavior interaction) to account for even less variance among supervisor ratings relative to self-ratings, in comparison to self relative to peer ratings.

The reasons that more extreme differences may exist between supervisor and self-ratings are varied. First, in addition to the factors that may lead to differences between peer and self-ratings stated above, one commonly cited weakness of supervisor ratings is that they exhibit more halo than either self or peer ratings (e.g., Lance, LaPointe, and Stewart, 1994). Halo refers to the notion that observers rate individuals high or low on some dimensions of performance, as a result of their performance on other, potentially independent dimensions of performance (Cascio, 1998). Cognitive researchers have attributed patterns of halo in ratings to observers’ tendencies to base their ratings on general impressions of an individual’s performance, rather than considering each performance dimension or behavior as a relatively distinct aspect of performance (Lord & Maher, 1991).

The use of such general impressions may be particularly widespread among supervisors due to the heavy cognitive load placed on them to rate multiple subordinates on multiple dimensions of performance. Indeed, in the context of rating individuals’ performance, a single supervisor is typically faced with making more ratings compared to a single peer (Greguras & Robie, 1998). To supervisors and peers alike, relying on general impressions serves as a
simplifying heuristic that lessens their cognitive load. Within the OCB literature, supervisors often rate multiple subordinates with regard to their engagement in multiple OCBs. Thus, to the extent that supervisors rely on general impressions of individuals’ organizational citizenship, one might expect that: (a) the individual facet contributes more to the variance in supervisor ratings of OCB relative to self-ratings and, to a lesser extent, peer ratings, and (b) the behavior-related sources of variation contribute less variance to supervisor ratings of OCB relative to self-ratings and, to a lesser extent, peer ratings.

**Peer-Supervisor Differences.** Differences between the variance architectures revealed by peer and supervisor ratings are expected to be much less severe relative to the differences expected between architectures revealed by self relative to supervisor ratings. Nevertheless, to the extent that confounding of the observer and group facet occurs in the case of supervisor ratings (and not peer ratings), one would expect the group facet to account for more variance in the supervisor ratings relative to peer ratings. Moreover, to the extent that supervisors must rate many more individuals than peers, behavior-related sources of variation may account for less variance in supervisor ratings relative to peer ratings (i.e., stemming from a potential greater reliance on general impressions as cited above; Lord & Maher, 1991).

**Summary.** Although several outcomes are expected from looking at the variance architectures of the OCB construct as revealed by different methods of measurement, it is important to qualify that even though the research cited above is suggestive of such differences between rating methods, the point of the VAA is not to confirm any of these explanations because it does not examine relationships among specific variables related to the root causes of these differences (e.g., observers cognitive loads, self-raters social desirability concerns etc.). What the VAA can provide is information that is either consistent or inconsistent with these claims based on the relative magnitudes of the facets of variance in question under each method. If the results
of this investigation via the VAA are consistent with these claims, then only subsequent research regarding specific variables related to these clams can determine whether they account for that variance in the facet of interest that is pertinent to the claims made above.

One potentially problematic aspect of making comparisons between architectures revealed by different methods of assessing OCB is that the contribution of the observer facet may be completely confounded with other VAA facets. For example, if a different supervisor (observer) provides ratings for all individuals in his/her workgroup (a common, though not universal practice in performance appraisal situations), separating out workgroup effects from observer effects is impossible. Therefore, any group effect that exists may be exacerbated due to bias stemming from how supervisors may be differentially using the scale to rate individuals in terms of OCB engagement. Such an observer effect is commonly referred to as a form of rater bias (e.g., leniency or severity) in the literature and, without alternative designs (e.g., multiple supervisors rating each group), it is impossible to separate such bias from the true group effects (Hoyt & Kerns, 1999; Scullen et al., 2000). Unfortunately, in some cases confounding is unavoidable. For example, in the case of self-ratings such a confounding of facet effects is unavoidable because there is only one observer for each individual (i.e., that given individual). As discussed earlier, when such confounding occurs caution should be taken in attributing variance due to the facet as solely due to one source or the other because it may be somewhat a reflection of both (e.g., group differences in OCB engagement and observers’ idiosyncratic rating tendencies).

Architectures of OCB through a Multitrait-Multimethod Lens

The fact that OCB is a multidimensional construct that is often assessed via a variety of methods makes it possible to examine the variance architectures underlying the OCB construct through a multitrait-multimethod (MTMM) lens and raises a number of interesting research
questions to be addressed. Campbell and Fiske’s (1959) multitrait-multimethod (MTMM) framework is typically used to address issues regarding the discriminant and convergent validity of constructs (or dimensions of a construct) and their measures. Within the MTMM framework researchers examine patterns of correlation among a set of “traits,” each assessed by a common set of “methods.” Convergent validity is evidenced by high monotrait, heteromethod correlations, whereas low heterotrait, monomethod correlations provide evidence of discriminant validity. What the VAA offers is an alternative set of values for populating such a MTMM matrix. Specifically, one can populate the cells an MTMM matrix with values that index the similarity of variance architectures underlying the given trait-method combinations as opposed simply correlations.

As discussed earlier, in comparing two or more constructs for similarity, one often approaches the issue in correlational terms, however the VAA offers multiple points of comparison. Specifically, one can populate an MTMM matrix with $ASI-RC$ and $ASI-RS$ values to get an idea of the similarity of the architectures underlying various OCB dimension-measure combinations and then examine architectures underlying such combinations for the specific sources of their similarity or dissimilarity (i.e., differences in the contribution of various facets within their architectures). Such multiple comparison points (points of similarity and dissimilarity) in constructs' architectures as revealed by the measures in question would arguably enable researchers to better pinpoint where the sources of convergent and discriminant validity lie and thus augment the MTMM approach which is based solely on correlational information drawn from a single-level of analysis.

In light of the observations made above, the present investigation examined two questions regarding the architectures underlying the dimensions of OCB as revealed by various methods of measurement. First, are the architectures underlying the same dimension of OCB as
revealed by different methods of measurement more similar than architectures of (a) different
dimensions of OCB as revealed by the same method of measurement and (b) different dimensions
of OCB as revealed by different methods of measurement? Such a finding would parallel findings
of high monotrait, heteromethod correlations in traditional MTMM models that are suggestive of
convergent validity. Second, are the architectures underlying different dimensions of OCB as
revealed by the same method of measurement less similar than architectures of (a) the same
dimension of OCB as revealed by different methods of measurement and (b) different dimensions
of OCB as revealed by different methods of measurement? Such a finding would parallel findings
of low heterotrait, monomethod correlations in traditional MTMM models that are suggestive of
discriminant validity. To the extent that such patterns of similarity are not apparent in the
architectures underlying the different dimensions OCB and their methods of measurement,
exploration of the specific sources of deviance from such patterns may readily be explored by
examining facet-level differences between the architectures underlying various dimension-
method combinations.

Summary

The present investigation relied on archival data from the OCB literature to provide (a) a
concrete example of the information that can result from the VAA, (b) answers to general types
of questions available from such information, (c) answers to research questions that are specific
to the OCB literature, and (d) an illustration of the challenges that may arise as one works through
the VAA research process. Table 5 summarizes the research questions that were examined in the
present investigation of the variance architectures underlying the OCB construct. The sections
that follow describe how OCB data was gathered for the present investigation and subsequently
analyzed to provide answers to the question summarized in Table 5.
Table 5. Summary of OCB-Related Research Questions

<table>
<thead>
<tr>
<th>OCB Research Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Where is the OCB literature currently lacking in terms of its ability to provide a basis for the full delineation of OCB's variance architecture?</td>
</tr>
<tr>
<td>- Does the OCB research literature contain data that are sampled across all VAA facets?</td>
</tr>
<tr>
<td>- Does the OCB research literature contain data that are coded in such a way that they allow for identification of the units of the VAA facets from which they are sampled?</td>
</tr>
<tr>
<td>- Is the OCB research literature characterized by measurement and sampling designs that allow for the unique estimation of each VAA facet’s contribution to variance in the OCB construct?</td>
</tr>
<tr>
<td>What is the contribution of each architectural facet to variance the OCB construct and what is the stability of those contributions across organizations?</td>
</tr>
<tr>
<td>- Does accounting for the multidimensionality of the OCB construct reduce the amount of variation attributable to behavior-related facets in architectures that ignore such multidimensionality?</td>
</tr>
<tr>
<td>- Are the architectures of the OCB dimensions nested within each higher-order OCB dimension (i.e., OCBI and OCBO) more similar to each other than the architectures of the OCB dimensions in the other higher-order OCB dimension?</td>
</tr>
<tr>
<td>How do the variance architectures of different dimensions of OCB compare to one another?</td>
</tr>
<tr>
<td>- Do individual, group, and organizational facets account for less variance among self-ratings compared both supervisor and peer ratings?</td>
</tr>
<tr>
<td>- Are these differences more salient when comparing self-ratings and supervisor ratings?</td>
</tr>
<tr>
<td>- Does the individual-by-behavior interaction, and indeed behavioral sources of variation in general, account for more variance among self-ratings compared to both supervisor and peer ratings?</td>
</tr>
<tr>
<td>- Are these differences more salient when comparing self-ratings to supervisor ratings?</td>
</tr>
<tr>
<td>- Are architectures of OCB revealed by peer and supervisor ratings generally more similar to each other than the architectures of OCB revealed by self-ratings?</td>
</tr>
<tr>
<td>Are the architectures underlying the same dimension of OCB as revealed by different methods of measurement more similar than architectures of (a) different dimensions of OCB as revealed by the same method of measurement, and (b) different dimensions of OCB as revealed by different methods of measurement (convergent evidence)?</td>
</tr>
<tr>
<td>Are the architectures underlying different dimensions of OCB as revealed by the same method of measurement less similar than architectures of (a) the same dimension of OCB as revealed by different methods of measurement, and (b) different dimensions of OCB as revealed by different methods of measurement (discriminant evidence)?</td>
</tr>
</tbody>
</table>

Method

The approach used to gather data for the present investigation can be viewed as similar to that of a meta-analysis, in that the data for this study was drawn from past studies that constitute the OCB literature. Unlike meta-analyses however, the data needed to conduct this study cannot simply be acquired from the information presented in the published works. For this study, the raw
data sets from these studies were needed. Thus, to acquire data for this investigation, raw data sets were sought from researchers who had gathered data on employees' engagement in OCB in previous studies. The process for identifying data sets for inclusion in this investigation is detailed below. In describing this process, the judgment calls that were made at each step of the data gathering process are made clear (Wanous, et al., 1989).

**Identification of Data Sets for Inclusion in the Present Study**

Given that the formal study of OCB did not begin until 1983 (Bateman & Organ, 1983; Smith et al., 1983), the search for articles published on OCB was limited to the years 1983-2000. Three steps were taken to identify studies that potentially gathered OCB data during this period. First, a literature search was conducted using three on-line databases: PsycInfo, ABI-Inform, and ISI's Social Science Citation Index. Keyword searches on the first two databases were conducted using the following words to generate an initial list of studies for possible inclusion in the present analysis: organizational citizenship behavior, OCB, citizenship behavior, organizational citizenship, prosocial behavior, pro-social behavior, helping behavior, altruistic behavior, extrarole behavior, extra-role behavior, contextual performance, and organizational spontaneity. All of these concepts have been associated with OCB in the academic literature.

The second step in the literature search process was to conduct a citation search on ISI's Social Citation Index for the two seminal articles of the OCB literature (e.g., Smith et al., 1983; Bateman & Organ, 1983). All articles uncovered in this were added to the list of articles identified in the first step of the literature search process. The third step in the search process was to add the studies examined in the Organ and Ryan’s (1995) meta-analysis of the OCB-antecedent literature. Upon completing these three steps, the resulting list was searched for duplicate citations and duplicates were removed.
Establishing Criteria for Inclusion

The next step in the data gathering process was to establish criteria for including studies in the present investigation that were identified by the initial search process described above. These criteria were based on the ability of the study to provide data that would be useful for delineating the variance architecture in the OCB construct.

The first criterion for inclusion was that the study needed to focus on individuals’ behavior in the context of organizations, rather than just individuals’ engagement in helping behavior divorced from the organizational context. This information was ascertained by reviewing the abstracts of all the studies under consideration. The elimination of studies based on their lack of focus on behavior in the organizational contexts was an ideal place to start because the majority of initial "hits" dealt with studies of helping behavior in everyday life stemming the social psychology literature. Reviewing the abstracts of the initial hits provided an efficient means for eliminating a vast majority of studies that were irrelevant to the present investigation.

The second criterion for inclusion was that the study needed to assess individuals’ engagement in OCB using a set of behaviors based on one of the commonly used scales in the OCB literature (e.g., Moorman and Blakely, 1995; Organ, 1994; Podsakoff et al., 1990; Smith et al., 1983; Williams & Anderson, 1991). Upon applying the first two criteria to the studies gathered in the initial literature search, the list of potential studies fell below 300.

The third criterion for inclusion necessitated that the study collected data on individuals' engagement in OCB via a field study conducted in an organizational setting. Although the first criterion limited studies to those focused on individuals’ organizationally relevant behavior, it did not eliminate those studies that examined such behavior in laboratory settings. Thus, all lab
studies regarding OCB (e.g., those studies that used undergraduate students participating in experiments) were eliminated from the sample of studies.

The fourth criterion for inclusion was that the study in question needed to assess OCB for *multiple* individuals’ who worked for the organization(s) examined. For example, studies that examined only single individuals’ within a variety of organizations (e.g., studies that surveyed MBA students from night classes who were each employed in different organizations) were eliminated from the sample. Such studies were excluded because the organization, group, and individual effects are completely confounded in such studies.

Upon completing this process, 100 data sets were identified for inclusion in the present study. Through personal contacts established while attempting to acquire data from these studies, nine other potential data sets were identified for inclusion. Thus, the sampling frame of data sets for this study grew to 109. Seventy-eight researchers maintained these data sets (i.e., first authors, or primary study contacts). Citations for the sources of 107 of these data sets are denoted with asterisks in the references section of this manuscript. Note that two of the data sets used in the present study were unpublished and have yet to be formally presented in their entirety.

*Procedures for Acquiring Raw Data Sets*

Gathering data for this study required personally contacting each of the 78 researchers who maintained the aforementioned 109 data sets. To contact these researchers, a search for the contact information (i.e., e-mail address and office phone number) for the first authors of each of the studies in which these data sets appeared was conducted. Contact information was obtained for 73 of these researchers (93.6%) via searches of professional membership directories, university department web sites where the researcher was employed, and personal acquaintances. Of the five researchers for whom no contact information was available, one was deceased and information regarding the last four could not be located via the above means.
The 73 researchers for whom contact information was available were contacted via e-mail to (a) request OCB data from their study (or studies) and (b) any other OCB data sets they might have had that fit the description of the data I was seeking for this study. A general version of the letter e-mailed to the researchers can be found in Appendix D. This initial request described (a) the data I was seeking, (b) a brief description of how it would be used in this study, and (c) why their data would be helpful for this study's successful completion. If no response was received from researchers after 14 days from the mailing of initial requests, a follow-up e-mail was sent. A general version of this follow-up letter can be found in Appendix E. If researchers did not respond as of 1 month after the initial mailing, attempts were made to contact them via telephone at their places of employment. If researchers could not be reached directly by phone, messages were left reflecting the content of the follow-up letter e-mail. If researchers did not respond to either e-mail or telephonic follow-ups, I contacted the main office of the department where they worked, acquired their office hours, and called the researcher directly when they were scheduled to be in their office. The section below describes the results of this data collection process and provides insight into areas where (a) difficulties in collecting raw data from researchers exist and (b) the current OCB literature fails to provide sufficient facet information to provide a full description of the variance architectures underlying the OCB construct.

Overview of Data Collection Results

The data collection phase of this project lasted approximately three months. Upon ending the data collection effort, I had received definitive answers (i.e., data provided or data denied) from 62 of the 73 researchers that I had contacted (84.9%) with regard to 90 of the 104 data sets I was seeking (86.7%). Table 6 summarizes results of the data collection process.

As noted in Table 6, 19 of the 73 researchers that were contacted provided data for the present investigation (26.0%). In total, 20 of the 104 data sets that were sought for inclusion in
Table 6. Results of the Data Collection Process

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Researchers</th>
<th>Data Sets</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>% of Total</td>
</tr>
<tr>
<td>Total Sampled</td>
<td>73</td>
<td>.</td>
</tr>
<tr>
<td>Responses</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Definitive&lt;sup&gt;a&lt;/sup&gt;</td>
<td>62</td>
<td>84.9</td>
</tr>
<tr>
<td>No Response</td>
<td>6</td>
<td>8.2</td>
</tr>
<tr>
<td>Pending&lt;sup&gt;b&lt;/sup&gt;</td>
<td>5</td>
<td>6.8</td>
</tr>
</tbody>
</table>

Breakdown of Definitive Responses

| Provided Data         | 19          | 26.0      | 20         | 19.2       |
| Denied Data           |             |           |            |            |
| Could not access data<sup>d</sup> | 2          | 2.7       | 2          | 1.9        |
| Could not find data   | 15          | 20.5      | 30         | 28.8       |
| No longer had data<sup>e</sup> | 10         | 13.7      | 10         | 9.6        |
| No time<sup>f</sup>   | 10          | 13.7      | 14         | 13.5       |
| Philosophical issues<sup>g</sup> | 3          | 4.1       | 6          | 5.8        |
| Prohibited to share data<sup>h</sup> | 1          | 1.4       | 1          | 1.0        |
| Other Reason<sup>i</sup> | 4          | 5.5       | 7          | 6.7        |

Note. <sup>a</sup> "Definitive" indicates that the researcher provided a definitive response to the request for his or her data (i.e., he or she either provided data, or denied the request). <sup>b</sup> "Pending" indicates researchers who responded positively to the request, but have yet to provide data. <sup>c</sup> Some of the 45 researchers who denied requests for data did supply other data sets. Thus, the 19 researchers who provided data and the 45 researchers who denied data do not represent two distinct sets of researchers. <sup>d</sup> "Could not Access Data" indicates researchers who denied the request for data because it was stored on a medium they no longer supported (e.g., 5.25 inch floppy disks). <sup>e</sup> "No Longer had data" indicates researchers who denied the request for data because they had disposed of it. <sup>f</sup> "No time" indicates researchers who denied the request for data because they felt they did not have time to fulfill it. <sup>g</sup> "Philosophical issues" indicates researchers who denied the request for data due to philosophical concerns they had regarding this project (e.g., “one should collect their own data for a dissertation”). <sup>h</sup> "Prohibited to Share Data" indicates researchers who denied the request for data because its proprietary nature prevented them from sharing it. <sup>i</sup> "Other Reason" indicates researchers who denied the request for data for reasons other than the categories listed in this table.

This study were obtained (19.2%). Only six researchers failed to respond to multiple requests for their data and as of this writing, five researchers are still considering my request for data. One of the interesting pieces of information arising from this data collection process were the reasons
offered by researchers with regard to why they denied requests for their data. In total, researchers declined to share 70 of the 104 data sets I sought for inclusion in this study. The most common reasons why researchers denied the request for data was that they either (a) could not find it, (b) had no time to pull the data together, or (c) had discarded it altogether. These three reasons alone accounted for the denial of 54 of the 104 data sets (51.9%) sought for inclusion in this study (77.1% of all denials). Given that the median year of publication for the studies that I was denied access to was 1995 ($SD = 3.73$), compared to the median year of publication for studies that contained data I acquired was 1998 ($SD = 1.57$), perhaps the prevalence of first and third most common reason cited above should not be surprising.

*An Overview of the Data Obtained.* In outlining the criteria that delineated the set of studies to sample for this study, no criterion was established for requiring a particular level of study quality. Nevertheless, I did establish a filter based on the quality of the data researchers provided. Although researchers provided 20 data sets for this investigation, only 9 of them were usable. The primary reasons why data sets were not usable is that they either (a) failed to identify the units of too many VAA facets (8 data sets) or (b) did not contain sufficient behavioral content with regard to any single dimension of OCB (3 data sets). For example, a data set may have distinguished between individuals, but may not contain any information with regard to individuals’ group membership (i.e., no group unit identifiers), or which observers rated which individuals with regard to their OCB engagement (i.e., no observer unit identifiers). Such a lack of unit identifiers made it impossible to decompose variance in the OCB construct as a function of the given facets. Although a lack of unit identifiers for certain facets in any given study is likely (e.g., those facets that are not central to the theory being examined in particular study), the seven data sets excluded from the present study based on their lack of facet identifiers contained
an extreme lack of facet information. Attempts to obtain the unit identifiers that were missing in these data sets directly from the authors of these studies were unsuccessful.

The other primary reason why data sets were excluded from this study stemmed from their lack of behavioral content with regard to any of the dimensions commonly studied as part of the OCB construct. Specifically, three data sets were excluded because they (a) only contained ratings of individuals’ OCB engagement on behaviors that were outside of core set of behaviors typically examined in the OCB literature (see Appendix F for a listing of behaviors typically examined in the OCB literature and that were used in the present study), or (b) contained too few behaviors per dimension to allow any behavioral-level variation to emerge (e.g., one behavior in a given OCB dimension). For these reasons, the three data sets referenced above were excluded from subsequent analyses.

Whereas the latter exclusion of studies may be indicative of a weakness of the VAA in general (i.e., its reliance on the literature’s reuse of similar measures), the exclusion of the former studies (i.e., those with insufficient facet coding) suggests that the present study only be viewed as a first attempt (i.e. not definitive) to provide a delineation of the variance architectures underlying the OCB construct. Indeed, by excluding over half of the data gathered for the present study, as well as failing to obtain 80.8% of the data sets sought, a full delineation of the variance architectures underlying the OCB construct was limited in the present study. Table 7 contains a summary of the data that was used to delineate variance architectures underlying OCB.

As presented in Table 7, ratings of OCB engagement in these data sets were primarily provided by supervisors (8 of 9 data sets), with only three data sets containing self-ratings, and one data set containing peer ratings. 21 Only three data sets contained ratings of OCB engagement from multiple rating sources (i.e., 3, 7, and 8). Given that I examined architectures of OCB as revealed by different methods of measurements (i.e., self, peer, and supervisor ratings),
Table 7. Data Sets Examined in the Present Study

<table>
<thead>
<tr>
<th>Data Set</th>
<th>OCB Dimensions(^a)</th>
<th>Method(^b)</th>
<th>Units Sampled per Facet</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cultures</td>
<td>Organizations</td>
<td>Groups</td>
</tr>
<tr>
<td>1</td>
<td>Alt, CV, Con, Court, Sport</td>
<td>Sup</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>Alt, CV, Con, Court, Sport</td>
<td>Sup</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>Alt, Court, Sport</td>
<td>Sup / Self</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>Alt, CV, Sport</td>
<td>Self</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>Alt, Con</td>
<td>Sup</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>Alt, CV, Con, Court, Sport</td>
<td>Sup</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>Alt, CV, Court</td>
<td>Sup / Self</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>Alt, CV</td>
<td>Sup / Peer</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>Alt</td>
<td>Sup</td>
<td>1</td>
</tr>
</tbody>
</table>

Unique Units Sampled Per Facet across Data Sets

<table>
<thead>
<tr>
<th></th>
<th>Cultures</th>
<th>Organizations</th>
<th>Groups</th>
<th>Jobs(^c)</th>
<th>Individuals</th>
<th>Observers(^d)</th>
<th>Behaviors</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>42</td>
<td>706</td>
<td>.</td>
<td>4,147</td>
<td>2,356</td>
<td>25</td>
<td></td>
</tr>
</tbody>
</table>

Self-Ratings

<table>
<thead>
<tr>
<th></th>
<th>Alt, CV, Court</th>
<th>Cultures</th>
<th>Organizations</th>
<th>Groups</th>
<th>Jobs(^c)</th>
<th>Individuals</th>
<th>Observers(^d)</th>
<th>Behaviors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6</td>
<td>113</td>
<td>.</td>
<td>1,053</td>
<td>1,053</td>
<td>15</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Peer Ratings

<table>
<thead>
<tr>
<th></th>
<th>Alt, CV</th>
<th>Cultures</th>
<th>Organizations</th>
<th>Groups</th>
<th>Jobs(^c)</th>
<th>Individuals</th>
<th>Observers(^d)</th>
<th>Behaviors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>246</td>
<td>.</td>
<td>633</td>
<td>633</td>
<td>8</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Supervisor Ratings

<table>
<thead>
<tr>
<th></th>
<th>Alt, CV, Con, Court, Sport</th>
<th>Cultures</th>
<th>Organizations</th>
<th>Groups</th>
<th>Jobs(^c)</th>
<th>Individuals</th>
<th>Observers(^d)</th>
<th>Behaviors</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>39</td>
<td>671</td>
<td>.</td>
<td>3,788</td>
<td>671</td>
<td>25</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. \(^a\) Abbreviations for OCB dimensions examined in the data sets are as follows: Alt- Altruism, CV- Civic Virtue, Con- Conscientiousness, Court- Courtesy, and Sport- Sportsmanship. \(^b\) Abbreviations for methods of assessing OCB in the data sets are as follows: Sup- Supervisor ratings, Self- Self-ratings, and Peer- Peer ratings. \(^c\) Little or no job-related information was available in the data sets that could be tied directly to specific individuals sampled. \(^d\) Slashes indicate observers stem from two different types of ratings sources (e.g., supervisors / sampled). Slashes indicate observers stem from two different types of ratings sources (e.g., supervisors / peers). Also note, no data across different occasions (all data was cross-sectional in nature) or stimuli was gathered, thus the occasion and stimulus facet are not listed in this table, and were not included in any architectural design examined in the present study.
subsequent discussion focuses on the breadth of sampling of units across VAA facets for each of these measurement methods separately.

Organization Facet. Data sets were gathered in 42 organizations across a wide array of industries. Of the organizations sampled, nine were commercial retailers, eight were diversified manufacturers, seven were public utility companies, six were high technology firms, five were financial institutions, two were telecommunications firms, and the remaining five organizations were from assorted industries (e.g., health care, higher education, city government). Data based on supervisor ratings were available from 39 of the 42 organizations sampled, whereas data based on peer ratings and self-ratings were only available from 10 and 6 of the organizations sampled, respectively. Generally, the more units one samples from any given facet, the more precise the estimate of that facet’s contribution to the total variance in a construct will be (Smith, 1981). Thus, estimates of the contribution of the organization facet and its associated interactions to the total variance in OCB will be more precise (i.e., tighter confidence intervals) for analyses based on supervisor ratings, relative to those revealed based on peer or self-ratings.

Culture Facet. There was very little variation in the sampling of units from the culture facet. Specifically, 41 of the 42 organizations were drawn from locations in the continental United States of America. If one assumes country can be used as a proxy for culture (e.g., Hofstede, 1980; Lam et al., 1999), such a lack of variation in the sampling of units across cultures prevents the inclusion of culture as a facet in the current study. Moreover, such a lack of sampling across cultures also prevents the current study from exploring the stability of variance architectures across different cultures.

Group Facet. The only information available in these studies with regard to units of the group facet occurred at the workgroup-level. In these studies, groups of employees were identified as those who worked together under the same supervisor. The total number of unique
groups identified by the present study was 706. As was the case with organizational units, the number of group units available for analyses based on supervisor, peer, and self-ratings differed. There were 671 groups available for analysis of architectures revealed by supervisor ratings ($Mdn_{within\,organization} = 10, SD = 15.38$), 246 available for analyses based on peer ratings ($Mdn_{within\,organization} = 15, SD = 20.61$), and 113 available for analyses based on self-ratings ($Mdn_{within\,organization} = 10, SD = 21.37$). Information on other group-level units, such as departments, was not readily available in these studies and thus was not examined as an architectural facet in the present study.

All workgroups examined in the present study were nested within organizations. Thus, for all architectures examined in the present study, the group facet was nested within the organization facet. As such, separating the contribution of the group facet and the group-by-organization interaction was not possible given the nested nature of the data. Consideration of OCB data collected under conditions where groups are crossed with organizations (e.g., teams of consultants or temporary employees working with many different organizations) will aide in future attempts to determine the unique contribution of the group facet and group-by-organization interaction.

**Individual Facet.** With regard to the individual facet, data regarding 4,147 individuals was available. As was the case with the organization and group facets, the majority of data on individuals stemmed from studies that employed supervisor ratings of OCB engagement. Specifically, there were 3,788 individuals available for analyses based on supervisor ratings ($Mdn_{within\,organization} = 60, SD = 114.8$), 633 available for analyses based on peer ratings ($Mdn_{within\,organization} = 38, SD = 54.9$), and 1,053 available for analyses based on self-ratings ($Mdn_{within\,organization} = 77, SD = 207.9$).

All individuals examined in the present study were nested within workgroups. Thus, for all architectures examined in the present study, the individual facet was nested within the group
facet. As such, separating the contribution of the individual facet and the individual-by-group interaction was not possible given the nested nature of the data. Future addition to OCB data collected under conditions where individuals are crossed with groups (e.g., matrix organizations) will aide in determining the unique contribution of the individual facet and individual-by-group interaction.

*Job Facet.* Little information was available about the jobs individuals held beyond what was revealed by their organizational membership. Specifically, in the vast majority of data sets gathered, authors of the original studies simply documented the general types of jobs individuals held in the organizations where they collected their data, but did not tie that information to individual records in the data file. For example, the two most common types of organizations examined were manufacturing firms and public utilities. The only distinction that could be made between individuals in terms of jobs in these organizations were that individuals within the manufacturing firms held a variety of manufacturing jobs and within the public utilities held a wide variety of jobs. Thus, in most data sets, specific job identifiers at the level of the job (e.g., individual A is a customer service representative, individual B is a bank teller), or at the level of the job family (e.g., individuals A-D are bankers, individuals E-H are human resource consultants) were not listed for each individual employee. Because no identifying variable in the data sets would allow one to reliably distinguish between the jobs or job families with which individuals were associated, the job facet was not included in this study’s analyses.

*Observer Facet.* Unlike the job facet, samples of units from the observer facet were quite sizable. Unfortunately, the observer facet was completely confounded with other VAA facets reviewed above. The facets that the observer facet was confounded with depended on what source of ratings was examined. For architectures revealed by supervisor ratings, the units of the observer facet were the supervisors themselves. Each supervisor provided ratings for every
individual in his or her workgroup and each supervisor supervised a different workgroup. Thus, the group facet and observer facets were completely confounded for architectures revealed by supervisor ratings with the current data. Specifically, the data does not allow one to distinguish between true group differences in OCB engagement (variation stemming from the group facet) and the idiosyncratic rating tendencies (e.g., differential use of the OCB rating scale) of individual supervisors (variation stemming from the observer facet). Alleviating such a confound among group and observer facets can be achieved by having multiple supervisors rate each group member’s engagement in OCB or by having supervisors rate multiple groups of individuals.

For self and peer ratings, a similar confounding occurs, only in this case the confound is between the individual facet and the observer facet. In the case of self-ratings, each individual provides ratings of his or her own engagement in OCB. Thus, the data does not allow one to distinguish between true individual differences in OCB engagement (variation stemming from the individual facet) and the idiosyncratic rating tendencies of individuals’ rating their own OCB engagement (variation stemming from the observer facet). Unfortunately, this confounding of individual and observer facets is a characteristic of self-ratings and cannot be alleviated through use of alternative measurement design or sampling strategies.

With regard to peer ratings, many individuals were rated by only one peer and typically the one peer was different for each individual being rated. Given the measurement design, it is not possible to distinguish between true individual differences in OCB engagement (variation stemming from the individual facet) and the idiosyncratic rating tendencies of the different peer raters (variation stemming from the observer facet). Alleviating the confound can be achieved by gathering multiple peers’ ratings of each individual’s engagement in OCB or having peers rate multiple individuals.
Behavior Facet. The last facet listed in Table 7 is the behavior facet. A total of 25 unique Organizational Citizenship Behavior items were used in the nine data sets examined (see Appendix F for the full list of items examined). However, not every study employed the same items in their administration of the OCB scale to observers. As the second column of Table 7 reveals, some studies provided more complete coverage of the OCB domain in terms of drawing items from each of its five factors than others. Moreover, peer-rating data were only available on the Altruism and Civic Virtue dimensions of OCB and self-rating data were only available on Altruism, Civic Virtue, and Courtesy. Only the supervisor-rating data allowed for the delineation of variance architectures underlying all five dimensions of OCB.

Summary. Based on the review of data gathered for the present study, it becomes readily apparent that delineating fully detailed architectures for the OCB construct was not possible in the present study. Culture, job, occasion, and stimulus facets were not included in any of the designs that are examined due to their unavailability in the current data. Furthermore, lack of self- and peer-rating data only allowed for the examination of architectures underlying some of the dimensions of OCB based on those rating sources. Lastly, several confounds stemming from the orientation of the facets to each other in the current data are present in the data (e.g., observer and group facet confound for supervisor ratings, observer and individual facet for self and peer ratings, nesting of individual, group, and organization facets). The orientation of the aforementioned facets to one another make it impossible to conclude that the variation resulting from one facet is solely due to variation across the units of that facet (i.e., it may stem from variation across the units of the facet it is confounded with). Although limitations on the architectures that can be examined in the present study are disconcerting, there is still value in examining architectures based solely on data that is available (e.g., the individual, group,
organization, and item facets) and attempting to address the research posed earlier based on such data.

Although the lack of data for many facets is unfortunate, the sample of OCB data appears to be quite representative of the population of existing OCB literature from which it was drawn. For example, organizations of many different sizes and industry, from a variety of locations across the U.S provided this data. Moreover, the OCB studies that were sampled contained similar proportions of supervisor ratings, peer ratings, and self-ratings relative to the frequency with which each of these rating types is generally used in the OCB literature (Podsakoff, et al., 2000). Nevertheless, while the sample I obtained in this study may be representative of the existing OCB literature, the samples of units obtained for each facet examined may not be entirely representative of the population of units on those facets, and thus the generalizability of these findings may be limited.

For example, 41 of the 42 organizations sampled in this investigation were located in the United States. Although one organization was located outside of the United States, the sample of organizations is not likely entirely representative of the population of all organizations across the world. In light of this fact, I am limited (to a certain extent) in the inferences, I can draw from the results of this study. For example, it is not is not possible to assess the degree to which findings of this study may remain stable across cultures (using nation where the organization is based as a proxy for culture). In light of these observations, care should be taken in interpreting the results of this study, with careful attention paid to the quality of sampling across each facet included in the architectures examined, and the potential impact that missing facets may have on inferences one attempts to draw from these results.
Laying out Architectures for OCB based on the Data Available

As illustrated in the previous section, the data made available for the present study limits the detail with which architectures underlying the OCB construct can be examined. Nevertheless, both highly detailed and more general architectures can provide valid descriptions of the dispersion of variance underlying a construct. Consider again the analogy of architectures underlying physical structures and the variance architectures underlying constructs. Descriptions of what a building looks like from a person standing on the street and descriptions of what a building looks like based on architectural blueprints would surely differ in their degree of detail. Nevertheless, both descriptions of the building are valid, one simply provides more detail on the structure of the building than the other, whereas the other provides more of a general overview of the structure of the building. For obtaining a broad overview of the dispersion of variance underlying a construct, a general description is useful. For achieving a more thorough description of the dispersion of variance underlying a construct, a more detailed exploration of a construct's variance architecture is required. The architectures examined in the present study vary in their degree of detail in an attempt to provide the benefits of (a) offering a general description of the dispersion of variance underlying the OCB construct, (b) a more detailed description of that dispersion of variance, and (c) determining the degree to which certain levels of detail are necessary to adequately describe that dispersion of variance.

The first architecture laid out in the present investigation was a design that is very common in the multilevel literature, namely decomposing variance in the OCB construct as a function of individual, group, and organization facets (Bryk & Raudenbush, 1992). The nature of this design is such that individuals (i) are nested within groups (g), which in turn are nested within organizations (o) (i.e., i:g:o). There are four sources of variance that can be estimated in this design (i.e., variance stemming from (a) individuals nested within groups, (b) groups nested
within organizations, (c) organizations, and (d) residual sources). The residual sources of variance in this design reflect any variance in the OCB construct that is not accounted for by the individual, group, and organization random facets. In this design, residual variance is completely confounded with variance arising from the individual-by-behavior interaction because multiple behaviors were rated for each individual sampled under this design.\(^{24}\) Estimating the parameters of this first general architecture provided a very general description of the dispersion of variance underlying the OCB construct. Parameters of this first architecture were estimated separately for (a) each rating source-by-OCB dimension combination for which OCB data was available (e.g., self-rated Altruism, peer-rated Civic Virtue, supervisor-rated Civic Virtue) and (b) each rating source across all behaviors where the sample of behaviors was not split by OCB dimension.

Although the general architectures described above provided gross overviews of the variance underlying the OCB construct, a goal of the VAA is to provide a very detailed, facet-rich description of the variance underlying a construct. Based on the data available, the most detailed architecture that could be estimated based on the available OCB data can be described as individuals, nested within groups, nested within organizations and crossed with behaviors (i.e., i:g:o x b). This design reveals seven potential sources of variance in individuals' engagement in OCB, namely, variance stemming from (a) organizations (e.g., the mean level of engagement in OCB is higher in some organizations than others), (b) groups nested within organizations (e.g., the mean level of engagement in OCB is higher in some groups than others within organizations), (c) individuals nested within groups (e.g., some individuals within groups engage in OCBs more than others), (d) behaviors (e.g., some OCBs are engaged in more than others), (d) the organization-by-behavior interaction (e.g., some OCBs are engaged in more in some organizations than others), (e) the group-by-behavior interaction (e.g., some OCBs are engaged in more in some groups within organizations than others), and (g) residual sources, which includes
the individual-by-behavior interaction (i.e., some OCBs are engaged in by some individuals more than others). Because there is only one observation per cell in this design, one cannot separate residual sources of variance and from variance stemming from the individual-by-behavior interaction term. Although, this confounding in the residual term represents a problem, there is little that can be done to uniquely estimate the behavior-by-employee effect unless one has multiple observations on each individual-by-behavior combination (e.g., repeated measurement occasions). As was the case with the first architecture examined (i.e., i:g:o), the parameters of this second architecture (i:g:o x b) were estimated separately for each rating source-by-dimension combination for which OCB data was available (e.g., self-rated Altruism, peer-rated Civic Virtue, supervisor-rated Civic Virtue).  

Because the ideal VAA *model* would incorporate as many facets as possible, efforts to examine the relative stability of the architectures focused solely on the i:g:o x b architecture introduced above (i.e., the more detailed architectures of OCB examined in the present study). Unfortunately, the data gathered for this study only allowed for the examination of stability across organizations. Lack of variation in the sampling of cultures prevented the examination of the stability of architectures across cultures. Furthermore, given that the earliest date that data sets for the current examination were gathered was 1997, it was not deemed worthwhile for examining the stability of architectures across different time periods.

Recall from the methodology described earlier that examining the stability of architectural facets required fitting a model for each organization separately. In doing so the organization and organization-by-behavior terms that were part of the i:g:o x b architecture described above drop out of the model. As such the design that was fit for each organization could be described as individuals nested within groups and crossed with behaviors (i.e., i:g x b). Such a design allowed for the examination of the stability of the contribution of five terms across
organizations, specifically the (a) individual facet, (b) group facet, (c) behavior facet, (d) group-by-behavior interaction, and (e) residual sources of variation (confounded with the individual-by-behavior interaction).

Summary. Table 8 provides a summary of the variance architectures examined in the present study and the data sets on which their parameter estimates were based. As Table 8 reveals, architectures could not be examined for all possible rating source-by-dimension combinations due to lack of self- and peer-rating data for some dimensions of OCB. In the next section, the general strategy for delineating the architectures summarized in Table 8 is briefly reviewed.

General Analysis Strategy

Before estimating the parameters of the variance architectures outlined in the previous section, several adjustments were made to the raw OCB data that was gathered. First, the OCB data from each study was standardized based on the range of the rating scale on which it was originally collected. Most measures of OCB use numerical rating scales ranging from either 1 to 5 or 1 to 7, indicating how frequently individuals engage in a given behavior. Because the 5-point scales were used more frequently, the ratings based on 7-point scales were scaled down to the 5-point scale (i.e., standardizing observations based on the rating scale range).26

The second adjustment made to the raw data was the elimination of cases that resulted in the redundancy of effects in the designs being estimated. Specifically, data from organizations that did not contain at least two workgroups were eliminated due to the redundancy of the organizational and group effects for such cases (i.e., one group nested in an organization does not allow for differentiation between the group and organization effects). Similarly, data from workgroups that did not contain at least two individuals were eliminated due to the redundancy of the group and individual effects for such cases. Failure to eliminate cases such as these has been
Table 8. Summary of Variance Architectures Examined in the Present Study

<table>
<thead>
<tr>
<th>Architectural Design</th>
<th>Target OCB Dimension</th>
<th>Supervisor Data Sets</th>
<th>Peer Data Sets</th>
<th>Self Data Sets</th>
</tr>
</thead>
<tbody>
<tr>
<td>i:g:o</td>
<td>All^b</td>
<td>Run 1 (1,2,3,5,6,7,8,9); Run 2 (1,2,6)</td>
<td>8</td>
<td>3,4,7</td>
</tr>
<tr>
<td>i:g:o x d^c</td>
<td>All</td>
<td>1,2,6</td>
<td>8</td>
<td>3,4,7</td>
</tr>
<tr>
<td>i:g:o</td>
<td>Altruism</td>
<td>1,2,3,5,6,7,8,9</td>
<td>8</td>
<td>3,4,7</td>
</tr>
<tr>
<td>i:g:o x b</td>
<td>Altruism</td>
<td>1,2,3,5,6,7,8,9</td>
<td>8</td>
<td>3,4,7</td>
</tr>
<tr>
<td>i:g x b^d</td>
<td>Altruism</td>
<td>1,2,3,5,6,7,8,9</td>
<td>8</td>
<td>3,4,7</td>
</tr>
<tr>
<td>i:g:o</td>
<td>Civic Virtue</td>
<td>1,2,6,7,8</td>
<td>8</td>
<td>4,7</td>
</tr>
<tr>
<td>i:g:o x b</td>
<td>Civic Virtue</td>
<td>1,2,6,7,8</td>
<td>8</td>
<td>4,7</td>
</tr>
<tr>
<td>i:g x b</td>
<td>Civic Virtue</td>
<td>1,2,6,7,8</td>
<td>8</td>
<td>4,7</td>
</tr>
<tr>
<td>i:g:o</td>
<td>Conscientiousness</td>
<td>1,2,5,6</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>i:g:o x b</td>
<td>Conscientiousness</td>
<td>1,2,5,6</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>i:g x b</td>
<td>Conscientiousness</td>
<td>1,2,5,6</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>i:g:o</td>
<td>Courtesy</td>
<td>1,2,3,6,7</td>
<td>.</td>
<td>3,4,7</td>
</tr>
<tr>
<td>i:g:o x b</td>
<td>Courtesy</td>
<td>1,2,3,6,7</td>
<td>.</td>
<td>3,4,7</td>
</tr>
<tr>
<td>i:g x b</td>
<td>Courtesy</td>
<td>1,2,3,6,7</td>
<td>.</td>
<td>3,4,7</td>
</tr>
<tr>
<td>i:g:o</td>
<td>Sportsmanship</td>
<td>1,2,3,6</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>i:g:o x b</td>
<td>Sportsmanship</td>
<td>1,2,3,6</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>i:g x b</td>
<td>Sportsmanship</td>
<td>1,2,3,6</td>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>

Note. Data set numbers correspond to the data sets listed in Table 7. ^a For supervisor ratings, parameters of the first architecture (i:g:o - All) were estimated based on two different data analysis runs due to the lack of computing power available for estimating the i:g:o x dimension architecture using all studies in Run 1. ^b “All” indicates that the given architecture contained all behaviors across the dimensions of OCB that were available for a given rating source. ^c This architecture was compared to the more general architecture (i:g:o, Run 2) to determine whether including dimension reduced the residual variance of the more general architecture. ^d This architecture (and all subsequent i:g x b architectures) were examined for the stability of their parameters across organizations.

linked to estimation problems when attempting to fit a GLMM to data (e.g., non-positive definite matrices; see Littell et al., 1996). Only a handful of cases from each rating source were eliminated due for these reasons (i.e., less than 1% of the data gathered). The facet sample sizes reported in Table 7 reflect the state of the data after the deletion of these cases occurred.
Upon cleaning the data in the manner described above, simple descriptive statistics were generated for the self, peer, and supervisor ratings, with regard to overall and dimensional OCB ratings. Given the normality assumptions that underlie REML-based estimates of variance components, distributions of the sample random effects (e.g., Brennan 1983) were estimated, and those distributions were subsequently evaluated against a normal distribution using skewness and kurtosis values. Appendix G fully details how sample estimates of the random effects stemming from to the units of each VAA facet were generated for the above normality checks.

Upon examining the data for potential departures from normality, the parameters of the variance architecture designs summarized in Table 8 were estimated using SAS Proc Mixed (Littell, et al., 1996). Specifically, SAS Proc Mixed was employed to generate restricted maximum likelihood (REML) estimates of the variance components associated with the facets and interaction terms of each architecture examined. SAS's Mixed procedure provides a flexible and powerful means for estimating variance components and other parameters of the GLMM using a variety of potential estimation methods (e.g., EMS, MIVQUE0, ML, REML), for any variety of random and mixed designs. Moreover, the Mixed procedure allows one to fit a variety of covariance structures for D and R, and allows one to compare the relative fit of models based on different covariance structures (Littell et al., 1996).

Upon estimating the variance components (i.e., the absolute contribution of each facet and interaction to total variance) of each facet and interaction term for each of the architectural designs, the estimates were summed across facets and interactions within each architecture to obtain an estimate of total variance for the given OCB dimension (or overall OCB construct). Each facet and interaction term's variance component was then divided by this total to arrive at an estimate of the proportion of variance accounted for by each facet and interaction term (i.e., the relative contribution of each facet and interaction to total variance).
Upon generating estimates of the absolute and relative contribution of each facet to total variance in the focal OCB dimension or construct, confidence intervals were formed around these estimates for each facet. Ninety-percent confidence intervals around variance components associated with each facet and interaction were calculated using a variation on Satterthwaite’s (1946) procedure as suggested by Brennan (1983). To form confidence intervals around the relative contribution of facets and interactions (i.e., proportions of total variance), several steps were taken. First, the proportion corresponding to a given facet was converted to a normal $z$ deviate using Fisher's transformation for normalizing distributions of intraclass correlations (Fisher, 1925). Ninety-percent confidence intervals were formed around the normal $z$ deviate corresponding to each proportion. The lower and upper bounds of this interval were formed by subtracting and adding (respectively) 1.65 times the standard error of the normal $z$ deviate (also obtained from Fisher's transformation). Upon obtaining these lower and upper bounds in terms of $z$, the bounds were converted back to proportions via Fisher's $z$ to $ICC$ transformation (Fisher, 1925). Appendix B provides further details on the methods used to form confidence intervals around both relative and absolute VAA facet contributions.

To examine the stability of architectures across organizations, a model with the i:g x b design was fitted for each organization separately. This allowed for the generation of organization-specific estimates of the relative contributions of the individual, group, behavior, behavior-by-group, and behavior-by-individual terms to the variance in OCB for each organization. Upon generating these estimates in each organization, all values were converted to normal $z$ deviates using Fisher's transformation cited earlier. The resulting $z$-values and their standard errors for each term (for each OCB dimension-rating method combination) were subsequently meta-analyzed across organizations via Hunter and Schmidt's (1990) methodology. By using such methods, estimates for the true variance of each of the facets...
across organizations were generated. Appendix C provides details on the methodology used in conducting these meta-analyses.

Upon delineating the variance architectures for the designs summarized in Table 8 using the strategies described above, the architectures for the various rating source-by-dimension combinations were compared using the methods described earlier. By making such comparisons, answers to the research questions surrounding the OCB construct raised earlier are provided.

Results

Descriptive Information and Normality Checks

In the sections that follow, descriptive information regarding individuals' engagement in OCB as revealed by self-ratings, peer ratings, and supervisor ratings is examined. The means and standard deviations of OCB engagement aggregated to four common levels of analysis (i.e., individual, group, organization, and behavior-levels) are presented in Tables 9 through 11. Also appearing in these tables are kurtosis and skewness values corresponding to each facet's distribution of sample random effects.

Self-Ratings

Table 9 presents descriptive information for self-ratings of OCB engagement. Based on self-ratings, Altruism and Courtesy OCB appear to be engaged in more by individuals relative to Civic Virtue OCB ($Cohen's \ d_{Altruism-Civic \ Virtue} = .81; \ Cohen's \ d_{Courtesy-Civic \ Virtue} = .72$). In terms of the normality of distributions of random effects sampled from each facet, the distributions do not appear to depart from normality enough to have substantial impact on the REML-based variance component estimation methods adopted in this study (Harville, 1977a; Marcoulides, 1990; Waternaux, 1976). In general, the distributions tend to be slightly negatively skewed (as would be expected in the case of performance ratings, particularly among self-raters; Murphy & Cleveland, 1995). Moreover the distributions appear to be slightly positively kurtotic (longer tails),
Table 9. Descriptives and Normality Checks for Self-Ratings

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Facet</th>
<th>n</th>
<th>$M_{raw}$</th>
<th>$SD_{raw}$</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>i: g:o</td>
<td>1,053</td>
<td>3.95</td>
<td>.62</td>
<td>-.51</td>
<td>.57</td>
<td></td>
</tr>
<tr>
<td>g:o</td>
<td>113</td>
<td>4.02</td>
<td>.31</td>
<td>.17</td>
<td>-.21</td>
<td></td>
</tr>
<tr>
<td>o</td>
<td>6</td>
<td>3.91</td>
<td>.10</td>
<td>-.12</td>
<td>.88</td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>b</td>
<td>15</td>
<td>3.85</td>
<td>.33</td>
<td>-.99</td>
<td>.72</td>
</tr>
<tr>
<td></td>
<td>g x b</td>
<td>1,088</td>
<td></td>
<td>-.31</td>
<td>3.16</td>
<td></td>
</tr>
<tr>
<td></td>
<td>o x b</td>
<td>70</td>
<td></td>
<td>-.13</td>
<td>5.87</td>
<td></td>
</tr>
<tr>
<td></td>
<td>i: g:o x b (res)</td>
<td>10,675</td>
<td></td>
<td>-.70</td>
<td>2.14</td>
<td></td>
</tr>
<tr>
<td>Altruism</td>
<td>i: g:o</td>
<td>443</td>
<td>3.47</td>
<td>.81</td>
<td>-.30</td>
<td>.25</td>
</tr>
<tr>
<td></td>
<td>g:o</td>
<td>47</td>
<td>3.54</td>
<td>.38</td>
<td>-.19</td>
<td>.39</td>
</tr>
<tr>
<td></td>
<td>o</td>
<td>5</td>
<td>3.61</td>
<td>.19</td>
<td>-1.13</td>
<td>.75</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>5</td>
<td>3.58</td>
<td>.38</td>
<td>.00</td>
<td>-.04</td>
</tr>
<tr>
<td></td>
<td>g x b</td>
<td>191</td>
<td></td>
<td>-.88</td>
<td>2.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>o x b</td>
<td>20</td>
<td></td>
<td>1.33</td>
<td>1.55</td>
<td></td>
</tr>
<tr>
<td></td>
<td>i: g:o x b (res)</td>
<td>1,843</td>
<td></td>
<td>-.36</td>
<td>.47</td>
<td></td>
</tr>
<tr>
<td>Civic Virtue</td>
<td>i: g:o</td>
<td>1,053</td>
<td>3.99</td>
<td>.68</td>
<td>-.55</td>
<td>.84</td>
</tr>
<tr>
<td></td>
<td>g:o</td>
<td>113</td>
<td>4.06</td>
<td>.33</td>
<td>.36</td>
<td>.03</td>
</tr>
<tr>
<td></td>
<td>o</td>
<td>6</td>
<td>4.07</td>
<td>.20</td>
<td>-.09</td>
<td>-1.39</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>5</td>
<td>3.99</td>
<td>.18</td>
<td>-.38</td>
<td>-2.96</td>
</tr>
<tr>
<td></td>
<td>g x b</td>
<td>434</td>
<td></td>
<td>.01</td>
<td>3.20</td>
<td></td>
</tr>
<tr>
<td></td>
<td>o x b</td>
<td>23</td>
<td></td>
<td>-1.95</td>
<td>5.19</td>
<td></td>
</tr>
<tr>
<td></td>
<td>i: g:o x b (res)</td>
<td>4,190</td>
<td></td>
<td>-.71</td>
<td>2.17</td>
<td></td>
</tr>
</tbody>
</table>

Note. $n$ indicates the number of units sampled from a facet. $M_{raw}$ indicates the mean of raw OCB ratings across units of a facet. $SD_{raw}$ indicates the standard deviation of raw OCB ratings across units of a facet. Under a normal distribution, the skewness and kurtosis statistics reported here have an expected value of 0.

particularly among the interaction effects.

Peer Ratings

Table 10 presents descriptive information for peer ratings of OCB engagement.
Table 10. Descriptives and Normality Checks for Peer Ratings

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Facet</th>
<th>n</th>
<th>$M_{raw}$</th>
<th>$SD_{raw}$</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>i:g:o</td>
<td>618</td>
<td>3.42</td>
<td>.83</td>
<td>-.24</td>
<td>.32</td>
<td></td>
</tr>
<tr>
<td>g:o</td>
<td>232</td>
<td>3.41</td>
<td>.62</td>
<td>-.36</td>
<td>-.01</td>
<td></td>
</tr>
<tr>
<td>o</td>
<td>10</td>
<td>3.37</td>
<td>.12</td>
<td>1.29</td>
<td>1.36</td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>b</td>
<td>8</td>
<td>3.40</td>
<td>.19</td>
<td>.64</td>
<td>-.83</td>
</tr>
<tr>
<td>g x b</td>
<td>1,781</td>
<td></td>
<td></td>
<td>-.16</td>
<td>1.01</td>
<td></td>
</tr>
<tr>
<td>o x b</td>
<td>80</td>
<td></td>
<td></td>
<td>-.98</td>
<td>2.89</td>
<td></td>
</tr>
<tr>
<td>i:g:o x b (res)</td>
<td>4,700</td>
<td></td>
<td></td>
<td>-.08</td>
<td>1.39</td>
<td></td>
</tr>
</tbody>
</table>

| Altruism     | i:g:o  | 613  | 3.35      | .98        | -.29     | .66      |
| g:o          | 231    | 3.34 | .72       | -.39       | -.12     |
| o             | 10     | 3.30 | .13       | .70        | 1.42     |
| Overall      | b      | 5    | 3.34      | .09        | .70      | .72      |
| g x b        | 1,110  |      |           | -.23       | .69      |
| o x b        | 50     |      |           | -.70       | 3.48     |
| i:g:o x b (res) | 2,943 |      |           | .32        | .12      |

| Civic Virtue | i:g:o  | 612  | 3.54      | .78        | -.14     | .07      |
| g:o          | 231    | 3.52 | .57       | -.34       | .08      |
| o             | 10     | 3.47 | .17       | .17        | -1.47    |
| Overall      | b      | 3    | 3.52      | .32        | -1.73    |          |
| g x b        | 671    |      |           | .03        | 1.02     |
| o x b        | 30     |      |           | -1.12      | 2.17     |
| i:g:o x b (res) | 1,757 |      |           | .26        | .39      |

Note. $n$ indicates the number of units sampled from a facet. $M_{raw}$ indicates the mean of raw OCB ratings across units of a facet. $SD_{raw}$ indicates the standard deviation of raw OCB ratings across units of a facet. Under a normal distribution, the skewness and kurtosis statistics reported here have an expected value of 0.

Unlike the sizable difference in individuals' self-ratings of Altruism OCB and Civic Virtue OCB, such a difference is not apparent among peer ratings. Indeed, based on peer ratings, individuals engage in Altruism OCB slightly less than Civic Virtue OCB ($Cohen's d_{Altruism-Civic Virtue} = -.21$). Comparing peer ratings to self-ratings revealed that Altruism OCB was rated substantially higher by self-raters relative to peer raters ($Cohen's d_{Altruism Self-Altruism Peer} = .88$), whereas Civic Virtue OCB ratings were not substantially different ($Cohen's d_{Civic Virtue Self-Civic Virtue Peer} = -.09$). Although such differences may arise from the fact that self-ratings and peer ratings were based on different
samples of individuals, to the extent that Altruism OCB is viewed as more socially desirable than Civic Virtue OCB, such findings are consistent with claims that self-raters may inflate ratings on dimensions of performance they view as most self-enhancing (Murphy & Cleveland, 1995). As with self-ratings, the normality of distributions of random effects sampled from each facet based on peer ratings appear to be fairly close to normal, and thus do not pose much of a threat to REML-based estimates of variance components for each facet. In general, the skewness and kurtosis values for peer ratings fall closer to 0 than those values for self-ratings (indicating more normality among peer-rating random effects).

**Supervisor Ratings**

Descriptives for supervisor ratings of OCB engagement are presented in Table 11. Supervisor ratings of Conscientiousness, Courtesy, and Altruism OCB suggest that individuals engage in these dimensions of OCB more than either Civic Virtue or Sportsmanship OCB. In comparison to self and peer ratings of Altruism and Civic Virtue, similarities and differences emerged. For example, the mean Civic Virtue rating at the individual-level was highly similar across all three rating methods. However, Altruism appeared to be rated higher by self-raters relative to both peer raters (as indicated above) and supervisors (Cohen's $d_{Altruism Self-Altruism Supervisor} = .37$). Differences between supervisor and self-ratings of Courtesy were similar to those found between those sources’ ratings of Altruism. Specifically, self-raters tended to provide higher mean ratings of Courtesy than supervisors ($Cohen's d_{Courtesy Self-Courtesy Supervisor} = .28$).

Although the skewness values associated with distributions of random effects sampled from each facet based on supervisor ratings appeared to be fairly close to normal, the kurtosis values for some isolated facets were high [e.g., the organization facet for Civic Virtue (8.17) and Sportsmanship (5.21)]. High kurtosis among the random effects from the organization facet may result in underestimates of its variance component because the number of organizations sampled
Table 11. Descriptives and Normality Checks for Supervisor Ratings

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Facet</th>
<th>$n$</th>
<th>$M_{raw}$</th>
<th>$SD_{raw}$</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>i:g:o</td>
<td>2,186</td>
<td>3.65</td>
<td>.72</td>
<td>-.37</td>
<td>1.01</td>
</tr>
<tr>
<td></td>
<td>g:o</td>
<td>406</td>
<td>3.67</td>
<td>.51</td>
<td>-.34</td>
<td>.85</td>
</tr>
<tr>
<td></td>
<td>o</td>
<td>39</td>
<td>3.66</td>
<td>.30</td>
<td>-.10</td>
<td>.60</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>25</td>
<td>3.70</td>
<td>.21</td>
<td>-.26</td>
<td>.19</td>
</tr>
<tr>
<td></td>
<td>g x b</td>
<td>5,269</td>
<td>.</td>
<td>.</td>
<td>-.44</td>
<td>2.57</td>
</tr>
<tr>
<td></td>
<td>o x b</td>
<td>486</td>
<td>.</td>
<td>.</td>
<td>-.46</td>
<td>3.68</td>
</tr>
<tr>
<td></td>
<td>i:g:o x b (res)</td>
<td>29,234</td>
<td>.</td>
<td>.</td>
<td>a</td>
<td>a</td>
</tr>
<tr>
<td>Altruism</td>
<td>i:g:o</td>
<td>2,186</td>
<td>3.76</td>
<td>.83</td>
<td>-.45</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>g:o</td>
<td>406</td>
<td>3.74</td>
<td>.58</td>
<td>-.46</td>
<td>.60</td>
</tr>
<tr>
<td></td>
<td>o</td>
<td>39</td>
<td>3.73</td>
<td>.34</td>
<td>-.13</td>
<td>.79</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>5</td>
<td>3.74</td>
<td>.12</td>
<td>-.05</td>
<td>.01</td>
</tr>
<tr>
<td></td>
<td>g x b</td>
<td>1,746</td>
<td>.</td>
<td>.</td>
<td>-.16</td>
<td>1.10</td>
</tr>
<tr>
<td></td>
<td>o x b</td>
<td>142</td>
<td>.</td>
<td>.</td>
<td>.19</td>
<td>1.73</td>
</tr>
<tr>
<td></td>
<td>i:g:o x b (res)</td>
<td>9,259</td>
<td>.</td>
<td>.</td>
<td>.06</td>
<td>.27</td>
</tr>
<tr>
<td>Civic Virtue</td>
<td>i:g:o</td>
<td>1,438</td>
<td>3.53</td>
<td>.87</td>
<td>-.23</td>
<td>.59</td>
</tr>
<tr>
<td></td>
<td>g:o</td>
<td>306</td>
<td>3.55</td>
<td>.62</td>
<td>-.14</td>
<td>.15</td>
</tr>
<tr>
<td></td>
<td>o</td>
<td>29</td>
<td>3.53</td>
<td>.36</td>
<td>-1.57</td>
<td>8.17</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>6</td>
<td>3.53</td>
<td>.20</td>
<td>-1.04</td>
<td>-.30</td>
</tr>
<tr>
<td></td>
<td>g x b</td>
<td>1,020</td>
<td>.</td>
<td>.</td>
<td>-.30</td>
<td>1.50</td>
</tr>
<tr>
<td></td>
<td>o x b</td>
<td>90</td>
<td>.</td>
<td>.</td>
<td>-.21</td>
<td>.22</td>
</tr>
<tr>
<td></td>
<td>i:g:o x b (res)</td>
<td>4,919</td>
<td>.</td>
<td>.</td>
<td>.07</td>
<td>.65</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>i:g:o</td>
<td>1,362</td>
<td>3.91</td>
<td>.85</td>
<td>-.45</td>
<td>.79</td>
</tr>
<tr>
<td></td>
<td>g:o</td>
<td>231</td>
<td>3.96</td>
<td>.57</td>
<td>-.93</td>
<td>2.67</td>
</tr>
<tr>
<td></td>
<td>o</td>
<td>24</td>
<td>3.89</td>
<td>.34</td>
<td>-1.06</td>
<td>2.34</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>5</td>
<td>3.94</td>
<td>.13</td>
<td>.52</td>
<td>-.97</td>
</tr>
<tr>
<td></td>
<td>g x b</td>
<td>861</td>
<td>.</td>
<td>.</td>
<td>-.50</td>
<td>1.96</td>
</tr>
<tr>
<td></td>
<td>o x b</td>
<td>81</td>
<td>.</td>
<td>.</td>
<td>.74</td>
<td>1.65</td>
</tr>
<tr>
<td></td>
<td>i:g:o x b (res)</td>
<td>5,127</td>
<td>.</td>
<td>.</td>
<td>-.05</td>
<td>1.27</td>
</tr>
</tbody>
</table>
Table 11- Continued

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Facet</th>
<th>n</th>
<th>$M_{raw}$</th>
<th>$SD_{raw}$</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Courtesy</td>
<td>i:g:o</td>
<td>1,221</td>
<td>3.77</td>
<td>.85</td>
<td>-.50</td>
<td>.82</td>
</tr>
<tr>
<td></td>
<td>g:o</td>
<td>199</td>
<td>3.82</td>
<td>.58</td>
<td>-.53</td>
<td>.77</td>
</tr>
<tr>
<td></td>
<td>o</td>
<td>20</td>
<td>3.75</td>
<td>.31</td>
<td>-.40</td>
<td>.89</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>5</td>
<td>3.78</td>
<td>.13</td>
<td>.57</td>
<td>-1.36</td>
</tr>
<tr>
<td></td>
<td>g x b</td>
<td>867</td>
<td>.</td>
<td>.</td>
<td>-.42</td>
<td>3.30</td>
</tr>
<tr>
<td></td>
<td>o x b</td>
<td>77</td>
<td>.</td>
<td>.</td>
<td>-.61</td>
<td>4.75</td>
</tr>
<tr>
<td></td>
<td>i:g:o x b (res)</td>
<td>5,336</td>
<td>.</td>
<td>.</td>
<td>.07</td>
<td>.49</td>
</tr>
<tr>
<td>Sportsmanship</td>
<td>i:g:o</td>
<td>1,136</td>
<td>3.49</td>
<td>1.13</td>
<td>-.36</td>
<td>.48</td>
</tr>
<tr>
<td></td>
<td>g:o</td>
<td>187</td>
<td>3.63</td>
<td>.85</td>
<td>-.18</td>
<td>1.73</td>
</tr>
<tr>
<td></td>
<td>o</td>
<td>18</td>
<td>3.63</td>
<td>.52</td>
<td>-1.95</td>
<td>5.21</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>3</td>
<td>3.50</td>
<td>.10</td>
<td>-.63</td>
<td>1.44</td>
</tr>
<tr>
<td></td>
<td>g x b</td>
<td>722</td>
<td>.</td>
<td>.</td>
<td>.07</td>
<td>2.14</td>
</tr>
<tr>
<td></td>
<td>o x b</td>
<td>64</td>
<td>.</td>
<td>.</td>
<td>-1.15</td>
<td>2.80</td>
</tr>
<tr>
<td></td>
<td>i:g:o x b (res)</td>
<td>4,529</td>
<td>.</td>
<td>.</td>
<td>.00</td>
<td>.75</td>
</tr>
</tbody>
</table>

Note. n indicates the number of units sampled from each facet. $M_{raw}$ indicates the means of raw OCB ratings across units of each facet. $SD_{raw}$ indicates the standard deviation of the mean raw OCB ratings across units of each facet. Under a normal distribution, the skewness and kurtosis statistics reported here have an expected value of 0.

from these facets is relatively small [i.e., 29 (Civic Virtue) and 18 (Sportsmanship); Tabachnick & Fidell, 1996]. In general, sampling more units from a facet that exhibits non-normal kurtosis may alleviate this problem. According to Waternaux (1976), underestimates of variance resulting from positive kurtosis begin to disappear with samples of 200 or more. Future efforts should be directed towards gathering data on facets where non-normality in distributions of their random effects may be evident. Sampling more units from such facets would help ensure that more precise, unbiased estimates of their variance components are obtained.30

Basic Variance Overview

Table 12 summarizes the results of delineating the first two variance architectures described in Table 8 for each of the three rating methods. In these first two architectures, OCBs from all dimensions were considered together in a single analysis. The purpose of examining the first architecture was to get a basic overview of the variance in OCB as it is distributed across
Table 12. Basic Variance Overview by Rating Method

<table>
<thead>
<tr>
<th>Design / Facet</th>
<th>Self-Ratings</th>
<th>Peer Ratings</th>
<th>Supervisor Ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Absolute</td>
<td>Relative</td>
<td>Absolute</td>
</tr>
<tr>
<td></td>
<td>$\sigma^2$</td>
<td>$\sigma^2$</td>
<td>$p_\sigma^2$</td>
</tr>
<tr>
<td></td>
<td>90% L U</td>
<td>90% L U</td>
<td>90% L U</td>
</tr>
<tr>
<td>i:g:o (without Dimension)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>i:g:o</td>
<td>.274 .250 .302</td>
<td>.262 .233 .292</td>
<td>.444 .389 .515</td>
</tr>
<tr>
<td>g:o</td>
<td>.017 .010 .045</td>
<td>.017 .009 .027</td>
<td>.159 .116 .250</td>
</tr>
<tr>
<td>o</td>
<td>.014 .006 .132</td>
<td>.013 .003 .044</td>
<td>.000 .000 .000</td>
</tr>
<tr>
<td>i:g:o x b (res)</td>
<td>.741 .723 .758</td>
<td>.708 .695 .721</td>
<td>.613 .591 .635</td>
</tr>
<tr>
<td>i:g:o x d (with Dimension)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>i:g:o</td>
<td>.244 . . . .221</td>
<td>. . . . .377</td>
<td>. . . . .315</td>
</tr>
<tr>
<td>g:o</td>
<td>.018 . . . .016</td>
<td>. . . . .121</td>
<td>. . . . .101</td>
</tr>
<tr>
<td>o</td>
<td>.000 . . . .000</td>
<td>. . . . .000</td>
<td>. . . . .000</td>
</tr>
<tr>
<td>d</td>
<td>.083 . . . .075</td>
<td>. . . . .013</td>
<td>. . . . .011</td>
</tr>
<tr>
<td>i x d</td>
<td>.051 . . . .046</td>
<td>. . . . .113</td>
<td>. . . . .094</td>
</tr>
<tr>
<td>g x d</td>
<td>.000 . . . .000</td>
<td>. . . . .051</td>
<td>. . . . .043</td>
</tr>
<tr>
<td>o x d</td>
<td>.016 . . . .015</td>
<td>. . . . .005</td>
<td>. . . . .004</td>
</tr>
<tr>
<td>i:g:o x b (res)</td>
<td>.690 . . . .626</td>
<td>. . . . .518</td>
<td>. . . . .432</td>
</tr>
</tbody>
</table>

**Note.** $\sigma^2$ indicates the estimated variance component for the given facet (absolute facet contribution). $\sigma^2$ 90% L and $\sigma^2$ 90% U indicate the lower and upper bounds (respectively) of the 90% confidence interval surrounding the $\sigma^2$ estimate for the given facet. $p_\sigma^2$ indicates the estimated proportion of total variance accounted for by the given facet (relative facet contribution) . $p_\sigma^2$ 90% L and $p_\sigma^2$ 90% U indicate the lower and upper bounds (respectively) of the 90% confidence interval surrounding the $p_\sigma^2$ estimate for the given facet.
levels of analysis commonly studied in the organizational behavior literature (i.e., individual, group, organization; Dansereau et al., 1984) as revealed by different rating methods. The purpose of examining the second architecture was to determine whether adding “dimension” to the model as a random factor resulted in an architecture with substantially better fit to the data relative to the first architecture (i.e., where dimension was not included).

Results of delineating the first architecture revealed that the largest source of variation in OCB engagement (for all rating methods) stemmed from the individual-by-behavior/residual term. Unfortunately, true variation due to the individual-by-behavior interaction is completely confounded with residual sources of variation because there is only one observation on each behavior for each individual. Given that these architectures examined behavior across all dimensions of OCB, a large portion of variance accounted for by the individual-by-behavior/residual term may be attributable to individuals differentially engaging in different dimensions of OCB (i.e., a true individual-by-dimension interaction); the second set of architectures were fitted to the data to address this possibility.

Results of fitting the second set of architectures revealed that adding a dimension factor and its related interactions to the first architecture significantly reduced both the relative and absolute contributions of the individual-by-behavior/residual term. Specifically, for self-ratings, peer ratings, and supervisor ratings, there were 11.6%, 14.3%, and 6.5% drops in the relative contribution of individual-by-behavior/residual term, and 6.9%, 15.5%, and 5.5% drops in the absolute contribution of this term. In all cases, architectures with dimension-related terms provided a better fit to the data than their corresponding architectures with dimension-related terms omitted. Table 13 summarizes the comparison of fit of architectures with and without dimension included in their designs.
Table 13. Comparing the Fit of Architectures with and without Dimension Information

<table>
<thead>
<tr>
<th>Rating Method</th>
<th>(-2 \text{ L } \Delta)</th>
<th>(\text{AIC } \Delta)</th>
<th>(\sigma^2\text{ residual } \Delta)</th>
<th>(p_{\sigma^2}\text{ residual } \Delta)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-Ratings</td>
<td>498.0†</td>
<td>-247.0</td>
<td>.051</td>
<td>.082</td>
</tr>
<tr>
<td>Peer Ratings</td>
<td>265.4†</td>
<td>-128.7</td>
<td>.095</td>
<td>.072</td>
</tr>
<tr>
<td>Supervisor Ratings</td>
<td>240.6†</td>
<td>-117.3</td>
<td>.041</td>
<td>.041</td>
</tr>
</tbody>
</table>

† \(p < .001\)

Note. \(-2 \text{ L } \Delta\) indicates the difference in \(-2\) residual log likelihood values for the first (i:g:o) and second (i:g:o x d) architectures. This value is \(\chi^2\) distributed with degrees of freedom equal to the difference in the number of parameters being estimated in each architecture (\(df = 4\) for all rating sources). Significant values indicate the architecture with dimension provided a better fit to the data. \(\text{AIC } \Delta\) indicates the difference in AIC information criterion values for the two architectures (0 indicates equal fit). Smaller AIC \(\Delta\) values indicate the architecture with dimension provided a better fit to the data. \(\sigma^2\text{ residual } \Delta\) indicates the difference between the variance components associated with the residual term for the architectures with and without dimension. \(p_{\sigma^2}\text{ residual } \Delta\) indicates the difference between the proportions of variance accounted for by the residual term for the architectures with and without dimension. All differences were calculated by taking values from architectures with dimension and subtracting them from values from architectures without dimension.

Although adding dimension to the first architecture reduced the contribution of the individual-by-behavior/residual term to total variance for all ratings methods, this term remained the largest source of variation in OCB ratings. Thus, there appears to be substantial variance that stems from individuals reacting differently to specific OCBs that may still be uncovered by looking at each dimension separately.

After, the individual-by-behavior/residual term, the largest source of variation across for all rating sources was the individual facet. Given that OCB is primarily studied as an individual-level construct, this finding should not be surprising. However, what is surprising is that the individual-by-behavior/residual term accounts for over twice as much variance as the individual facet for both self and supervisor ratings (about one-and-half times as much for peers). Typically, when studying OCB at the individual-level, specific OCBs are collapsed to obtain a more reliable measure of the OCB construct and thus such variance is not readily apparent. The fact that so much variance is attributable to this interaction term suggests that a substantial proportion of
variance in the OCB construct is being overlooked when one forms individual-level scales of OCB based on composites of OCB items. Previous research has revealed that this “residual” variance may be worth reconsidering in the context of examining OCB (e.g., Morrison, 1994; Putka & Vancouver, 1999).

Results of these first two sets of analyses provide partial support for the logic underlying the OCB research questions raised earlier. For example, based on the performance appraisal literature, one might expect that the individual, group, and organization facets would account for less total variance among self-ratings relative to both peer and supervisor ratings. This expectation stems from findings in the performance appraisal literature that individuals tend to inflate their performance ratings, which would put a ceiling on variation stemming from their ratings at the individual-level and aggregates of their ratings at higher levels (e.g., due to self-raters only using the top portion of the rating scale). Based on the present analyses, the variance accounted for by the individual facet among self-ratings ($\sigma_{i,g,o}^2 = .274$) was significantly less than the variance accounted for by the individual facet among peer ratings ($\sigma_{i,g,o}^2 = .444$), but not supervisor ratings ($\sigma_{i,g,o}^2 = .290$). However, the group facet accounted for significantly less variance among self-ratings ($\sigma_{g,o}^2 = .017$) relative to both peer ($\sigma_{g,o}^2 = .159$) and supervisor ratings ($\sigma_{g,o}^2 = .094$). With regard to the organization facet, this claim was only partially supported, as the amount of variance accounted for by the organization facet based on self-ratings ($\sigma_{o}^2 = .014$) was significantly less than the variance accounted for by the organization facet among supervisor ratings ($\sigma_{o}^2 = .060$), but actually was significantly greater than the variance accounted for by the organization facet among peer ratings ($\sigma_{o}^2 = .000$).

Recall from arguments made earlier that the individual-by-behavior/residual term was expected to account for more of variation among self-ratings compared to peer and supervisor ratings. Based on the present analyses, this claim appears to be partially supported, as the
variance component for the individual-by-behavior/residual term based on self-ratings ($\sigma^2 = .741$) was significantly greater than the variance component for that term among peer ratings ($\sigma^2 = .613$), but not for that term among supervisor ratings ($\sigma^2 = .746$). Interestingly, the relative contribution of the individual-by-behavior/residual term was significantly greater for self-ratings ($p_{\sigma^2} = .708$) compared to its relative contribution for both peer ($p_{\sigma^2} = .504$) and supervisor ratings ($p_{\sigma^2} = .627$).

Summary. Although the above architectures provide a basic overview of the variance in the OCB construct, it is important to remember that OCB is a multidimensional construct. As such, it is important to move beyond these general architectures and address the OCB research questions alluded to above as they pertain to each specific dimension of OCB. In sections that follow, variance architectures underlying each dimension of OCB are delineated.

Delineating Variance Architectures Underlying Each Dimension of OCB

Delineating the architectures underlying each dimension of OCB provided the most comprehensive look at the variance architectures underlying the OCB construct that is possible based on the current data. The purpose of this section is to describe the variance architectures underlying each dimension of OCB separately. Upon describing these architectures in detail, subsequent sections will make explicit architecture- and facet-level comparisons of the architectures described in this section. Any comparisons that are made in this section will focus primarily on facet-level differences in architectures underlying different dimensions of OCB within rating methods.

The following section is broken down by rating method (i.e., self, peer, and supervisor ratings). Within each rating method section, three architectures were examined for each dimension of OCB for which data was available. The first architecture that was delineated for each dimension (i.e., $i:g:o$) ignored behavior and primarily served as a baseline model against
which the necessity of more complex architectures (i.e., one’s that included behavior as a facet) could be evaluated. The second architecture that was delineated for each dimension (i.e., i:g:o x b) explicitly included behavior as a facet and provided the most detailed description of a dimension’s variance architecture that was possible based on the current data. To test the necessity of including the behavior facet in each architecture, the first and second architectures were compared for their relative fit to the data. The stability of facet contributions across organizations was examined by delineating a third architecture (i:g x b) for each organization separately and aggregating results meta-analytically as described earlier. The results of these analyses allowed for a complete description of each dimension’s variance architecture (for the facets available) in terms of the magnitude and stability of its facets’ contributions (across organizations).

**Self-Ratings**

Table 14 presents the results of the comparative fit analyses for architectures with and without the behavior facet. Recall from earlier descriptions of the data that no self-ratings were obtained for either the Conscientiousness or Sportsmanship dimensions, thus only architectures underlying Altruism, Civic Virtue, and Courtesy were delineated.

**Table 14. Comparing the Fit of Architectures Based on Self-Ratings**

<table>
<thead>
<tr>
<th>Dimension</th>
<th>-2 Ln Δ</th>
<th>AIC Δ</th>
<th>$\sigma^2$ Residual Δ</th>
<th>$\rho^2$ Residual Δ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Altruism</td>
<td>451.8†</td>
<td>-223.8</td>
<td>.080</td>
<td>.093</td>
</tr>
<tr>
<td>Civic Virtue</td>
<td>236.7†</td>
<td>-117.3</td>
<td>.179</td>
<td>.136</td>
</tr>
<tr>
<td>Courtesy</td>
<td>461.1†</td>
<td>-230.5</td>
<td>.110</td>
<td>.161</td>
</tr>
</tbody>
</table>

*Note. df = 3 for all dimensions. All differences were calculated by taking values from architectures with behavior and subtracting them from values from architectures without behavior. † p < .001
As revealed by Table 14, the architectures including the behavior facet provided a significantly better fit to the data than the architectures that excluded the behavior facet. Specifically, including the behavior facet in the architecture along with its corresponding interactions (group-by-behavior and organization-by-behavior) resulted in 14.8%, 20.7%, 22.3% reductions in the relative contribution of residual variance to total variance for the Altruism, Civic Virtue, and Courtesy dimensions respectively, and 14.5%, 17.9% and 14.7% reductions in the absolute contribution of residual variance for these dimensions. These findings indicate that a simpler architecture composed solely of facets corresponding to typical levels-of-analysis in the organizational behavior literature (i.e., individuals, groups, and organizations) would not suffice for describing the variance underlying these dimensions of the OCB construct as revealed by self-ratings. Given these findings, Table 15 only presents parameter estimates for variance architectures where behavior was included as a facet.

Magnitude of Facet Contributions. The largest source of variance in OCB engagement as revealed by self-ratings was the individual-by-behavior/residual term. This term accounted for between 52% (Civic Virtue) and 56.8% (Courtesy) of the total variance in these dimensions of OCB. The next largest source of variance in each of these dimensions was the individual facet, which accounted for between 24.6% (Courtesy) and 35% (Altruism) of the variance in OCB engagement. A striking similarity among the architectures of the three dimensions examined was the lack of contribution of either group or organization facet. The contribution of the organization facet was estimated to be near zero for all three dimensions, whereas at most, the group facet accounted for only 2% of total variance (Altruism).

Another interesting aspect of the architectures underlying these dimensions is that the organization-by-behavior interaction consistently emerged as the third largest source of variation. This interaction accounted for between 5.6% (Altruism) and 16.7% (Courtesy) of the variance in
Table 15. Variance Architecture Summary for Self-Ratings

<table>
<thead>
<tr>
<th>Facet</th>
<th>Absolute</th>
<th>Relative</th>
<th>Facet Stability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\sigma^2$</td>
<td>$\sigma^2_{90%L}$</td>
<td>$\sigma^2_{90%U}$</td>
</tr>
<tr>
<td>Altruism</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>i:g:o</td>
<td>.309 .280 .343</td>
<td>.350 .308 .391</td>
<td>6 .291 .131</td>
</tr>
<tr>
<td>g:o</td>
<td>.017 .010 .049</td>
<td>.020 .005 .038</td>
<td>6 .013 .057</td>
</tr>
<tr>
<td>o</td>
<td>.000 .000 .000</td>
<td>.000 .000 .007</td>
<td>. . .</td>
</tr>
<tr>
<td>b</td>
<td>.034 .013 .765</td>
<td>.039 .007 .160</td>
<td>6 .145 .237</td>
</tr>
<tr>
<td>g x b</td>
<td>.002 .001 .616</td>
<td>.002 .000 .028</td>
<td>6 .009 .026</td>
</tr>
<tr>
<td>o x b</td>
<td>.049 .031 .099</td>
<td>.056 .030 .094</td>
<td>. . .</td>
</tr>
<tr>
<td>i:g:o x b (res)</td>
<td>.472 .453 .490</td>
<td>.534 .506 .561</td>
<td>6 .474 .122</td>
</tr>
<tr>
<td>Civic Virtue</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>i:g:o</td>
<td>.446 .385 .525</td>
<td>.335 .273 .398</td>
<td>5 .292 .131</td>
</tr>
<tr>
<td>g:o</td>
<td>.000 .000 .000</td>
<td>.000 .000 .023</td>
<td>5 .018 .057</td>
</tr>
<tr>
<td>o</td>
<td>.000 .000 .000</td>
<td>.000 .000 .017</td>
<td>. . .</td>
</tr>
<tr>
<td>b</td>
<td>.083 .030 1.775</td>
<td>.062 .010 .253</td>
<td>5 .130 .237</td>
</tr>
<tr>
<td>g x b</td>
<td>.000 .000 .000</td>
<td>.000 .000 .045</td>
<td>5 .011 .026</td>
</tr>
<tr>
<td>o x b</td>
<td>.109 .063 .254</td>
<td>.082 .040 .147</td>
<td>. . .</td>
</tr>
<tr>
<td>i:g:o x b (res)</td>
<td>.691 .647 .734</td>
<td>.520 .477 .560</td>
<td>5 .523 .122</td>
</tr>
<tr>
<td>Courtesy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>i:g:o</td>
<td>.275 .246 .313</td>
<td>.246 .204 .289</td>
<td>6 .249 .213</td>
</tr>
<tr>
<td>g:o</td>
<td>.020 .011 .055</td>
<td>.018 .003 .037</td>
<td>6 .016 .054</td>
</tr>
<tr>
<td>o</td>
<td>.000 .000 .000</td>
<td>.000 .000 .008</td>
<td>. . .</td>
</tr>
<tr>
<td>b</td>
<td>.000 .000 .000</td>
<td>.000 .000 .010</td>
<td>6 .148 .294</td>
</tr>
<tr>
<td>g x b</td>
<td>.000 .000 .000</td>
<td>.000 .000 .027</td>
<td>6 .031 .092</td>
</tr>
<tr>
<td>o x b</td>
<td>.187 .122 .363</td>
<td>.167 .104 .254</td>
<td>. . .</td>
</tr>
<tr>
<td>i:g:o x b (res)</td>
<td>.636 .610 .663</td>
<td>.568 .541 .595</td>
<td>6 .548 .254</td>
</tr>
</tbody>
</table>

Note. * Value falls below Hunter-Schmidt’s 75% cut-off. k indicates the number of organizations across which stability was examined. M $p^2_\sigma$ indicates the mean estimated proportion of total variance accounted for by a given facet averaged across organizations. ISDR $p^2_\sigma$ indicates the raw inter-standard deviation range of the $p^2_\sigma$ estimates across organizations. % Var Artifact indicates the proportion of raw variance in $p^2_\sigma$ estimates across organizations that can be accounted for by sampling error.
individuals’ OCB engagement. This finding suggests that organizations may hold different policies that serve to differentially reinforce individuals’ engagement in various OCBs. Thus, depending on the organization, some OCBs may be engaged in by individuals more than other OCBs because such behaviors are being differentially rewarded, either formally or informally. Another explanation for this finding may be that the culture or norms espoused within the organization may indicate some OCBs are more socially desirable than others, thus potentially leading individuals’ to inflate their ratings on some OCBs more than others, to make them appear in a more favorable light given an organization’s specific culture (Feldman, 1984). Either of these explanations can be fodder for future research in the OCB literature.

**Stability of Facet Contributions.** The ability of this study to provide an adequate assessment of the stability of facet contributions across organizations is limited due to the small number of organizations for which self-ratings were available. Thus, the findings presented in Table 15 should be interpreted with caution as data were only sampled from six organizations. Given this qualifying statement, the findings presented in Table 15 suggest that nearly all of the variance in estimates of relative facet contributions across organizations can be attributed to sampling error. Of all the facets and dimensions examined, only the individual-by-behavior/residual term for Courtesy exhibited signs that substantial true variation may exist in facet contributions across organizations. This term accounted for between 35.9% and 65.7% of total variance in Courtesy across organizations ($M_p \sigma^2 = .548, ISDR p_\sigma^2 = .254$). Although sampling error only accounted for 73% of the variance in estimates of the relative contribution of the individual-by-behavior/residual term across organizations for the Courtesy dimension, a $Q$-test of homogeneity revealed no significant difference between the raw and error variance estimates for this term’s contribution across organizations ($Q = 8.22, ns$).
Peer Ratings

Table 16 presents the results of the comparative fit analyses for architectures with and without the behavior facet. Recall from earlier descriptions of the data that no peer ratings were obtained for the Conscientiousness, Courtesy, and Sportsmanship dimensions, thus only architectures underlying Altruism and Civic Virtue were delineated.

Table 16. Comparing the Fit of Architectures based on Peer Ratings

<table>
<thead>
<tr>
<th>Dimension</th>
<th>(-2 , L \Delta)</th>
<th>AIC (\Delta)</th>
<th>(\sigma^2) Residual (\Delta)</th>
<th>(p^2) Residual (\Delta)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Altruism</td>
<td>41.6†</td>
<td>-17.8</td>
<td>.050</td>
<td>.040</td>
</tr>
<tr>
<td>Civic Virtue</td>
<td>179.7†</td>
<td>-86.8</td>
<td>.150</td>
<td>.166</td>
</tr>
</tbody>
</table>

Note. df = 3 for all dimensions. All differences were calculated by taking values from architectures with behavior and subtracting them from values from architectures without behavior. † \(p < .001\)

As revealed by Table 16, the architectures including the behavior facet provided a significantly better fit to the data than architectures that excluded the behavior facet. Specifically, including the behavior facet in the architecture along with its corresponding interactions (group-by-behavior and organization-by-behavior) resulted in 12.1% and 25.2% reductions in the relative contribution of residual variance to total variance for Altruism and Civic Virtue respectively, and 11.6% and 21.4% in the absolute contribution of residual variance for these dimensions. As was the case with the self-ratings, these findings indicate that an architecture that ignores behavior-related variation would not suffice for describing the variance underlying these dimensions of the OCB construct. Given these findings, Table 17 only presents parameter estimates for variance architectures where behavior was included as a facet.

Magnitude of Facet Contributions. Unlike the variance architectures underlying different dimensions of OCB as revealed by self-ratings, the variance architectures of Altruism and Civic
### Table 17. Variance Architecture Summary for Peer Ratings

<table>
<thead>
<tr>
<th>Facet</th>
<th>Facet Contributions</th>
<th>Facet Stability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Absolute</td>
<td>Relative</td>
</tr>
<tr>
<td></td>
<td>( \sigma^2 )</td>
<td>( \sigma^2 )</td>
</tr>
<tr>
<td></td>
<td>90% L</td>
<td>90% U</td>
</tr>
<tr>
<td></td>
<td>( \rho_{\sigma^2} )</td>
<td>( ISDR )</td>
</tr>
<tr>
<td></td>
<td>( \rho_{\sigma^2} )</td>
<td></td>
</tr>
</tbody>
</table>

#### Altruism

<table>
<thead>
<tr>
<th>facet</th>
<th>( \sigma^2 )</th>
<th>( \sigma^2 )</th>
<th>( \rho_{\sigma^2} )</th>
<th>( \rho_{\sigma^2} )</th>
<th>( k )</th>
<th>( M_{\sigma^2} )</th>
<th>% Var</th>
<th>Artifact</th>
</tr>
</thead>
<tbody>
<tr>
<td>i:g:o</td>
<td>.670</td>
<td>.589</td>
<td>.771</td>
<td>.517</td>
<td>.481</td>
<td>.551</td>
<td>.448</td>
<td>.375</td>
</tr>
<tr>
<td>g:o</td>
<td>.196</td>
<td>.135</td>
<td>.319</td>
<td>.151</td>
<td>.120</td>
<td>.185</td>
<td>.166</td>
<td>.316</td>
</tr>
<tr>
<td>o</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.005</td>
<td>.006</td>
<td>.019</td>
</tr>
<tr>
<td>b</td>
<td>.005</td>
<td>.002</td>
<td>.112</td>
<td>.003</td>
<td>.000</td>
<td>.019</td>
<td>.006</td>
<td>.019</td>
</tr>
<tr>
<td>g x b</td>
<td>.047</td>
<td>.034</td>
<td>.073</td>
<td>.036</td>
<td>.000</td>
<td>.074</td>
<td>.033</td>
<td>.046</td>
</tr>
<tr>
<td>o x b</td>
<td>.000</td>
<td>.000</td>
<td>.120</td>
<td>.000</td>
<td>.000</td>
<td>.009</td>
<td>.000</td>
<td>.009</td>
</tr>
<tr>
<td>i:g:o x b (res)</td>
<td>.379</td>
<td>.356</td>
<td>.403</td>
<td>.293</td>
<td>.262</td>
<td>.323</td>
<td>.322</td>
<td>.227</td>
</tr>
</tbody>
</table>

#### Civic Virtue

<table>
<thead>
<tr>
<th>facet</th>
<th>( \sigma^2 )</th>
<th>( \sigma^2 )</th>
<th>( \rho_{\sigma^2} )</th>
<th>( \rho_{\sigma^2} )</th>
<th>( k )</th>
<th>( M_{\sigma^2} )</th>
<th>% Var</th>
<th>Artifact</th>
</tr>
</thead>
<tbody>
<tr>
<td>i:g:o</td>
<td>.274</td>
<td>.226</td>
<td>.345</td>
<td>.247</td>
<td>.199</td>
<td>.296</td>
<td>.248</td>
<td>.686</td>
</tr>
<tr>
<td>g:o</td>
<td>.102</td>
<td>.067</td>
<td>.193</td>
<td>.093</td>
<td>.060</td>
<td>.129</td>
<td>.095</td>
<td>.308</td>
</tr>
<tr>
<td>o</td>
<td>.011</td>
<td>.004</td>
<td>4.290</td>
<td>.010</td>
<td>.001</td>
<td>.031</td>
<td>.010</td>
<td>.031</td>
</tr>
<tr>
<td>b</td>
<td>.113</td>
<td>.056</td>
<td>54.759</td>
<td>.102</td>
<td>.009</td>
<td>.543</td>
<td>.080</td>
<td>.362</td>
</tr>
<tr>
<td>g x b</td>
<td>.052</td>
<td>.031</td>
<td>.135</td>
<td>.047</td>
<td>.000</td>
<td>.097</td>
<td>.054</td>
<td>.113</td>
</tr>
<tr>
<td>o x b</td>
<td>.009</td>
<td>.004</td>
<td>.257</td>
<td>.008</td>
<td>.000</td>
<td>.024</td>
<td>.000</td>
<td>.024</td>
</tr>
<tr>
<td>i:g:o x b (res)</td>
<td>.546</td>
<td>.498</td>
<td>.593</td>
<td>.493</td>
<td>.460</td>
<td>.525</td>
<td>.511</td>
<td>.443</td>
</tr>
</tbody>
</table>

*Note. a Value falls below Hunter-Schmidt’s 75% cut-off. b Observed and error variance significantly different based on Q-test of homogeneity, \( p < .05 \).*

Virtue revealed by peer ratings appear to have a number of dissimilarities. For example, the largest and second largest sources of variation in Altruism and Civic Virtue are not the same. Namely, the largest source of variation in Altruism is the individual facet (51.7% of total variance), whereas the largest source of variation in Civic Virtue is the individual-by-behavior interaction / residual term (49.3% of total variance). Furthermore, both the absolute \( (\sigma_{i:g:o}^2 = .670) \) and relative \( (\rho_{\sigma}^2 = .517) \) contributions of the individual facet for Altruism are significantly larger.
than their respective contributions for Civic Virtue ($\sigma_{\text{ego}}^2 = .274$; $p_{\sigma}^2 = .247$), and both the absolute ($\sigma_{\text{ego}}^2 = .546$) and relative ($p_{\sigma}^2 = .493$) contributions of the individual-by-behavior interaction/residual term for Civic Virtue are significantly larger than their respective contributions for Altruism ($\sigma_{\text{ego}}^2 = .379$; $p_{\sigma}^2 = .293$).

A second difference observed between the architectures of Altruism and Civic Virtue regarded the contribution of the group facet. Specifically, both the absolute and relative contributions of the group facet to variance in Altruism ($\sigma_{\text{ego}}^2 = .196$; $p_{\sigma}^2 = .151$) were significantly greater than their respective contributions to variance in Civic Virtue ($\sigma_{\text{ego}}^2 = .102$; $p_{\sigma}^2 = .093$).

Lastly, a third difference between the architectures of peer-rated Altruism and Civic Virtue stemmed from the behavior facet. Specifically, both the absolute and relative contributions of the behavior facet to variance in Civic Virtue ($\sigma_{\text{ego}}^2 = .113$; $p_{\sigma}^2 = .102$) were significantly greater than their respective contributions to variance in Altruism ($\sigma_{\text{ego}}^2 = .005$; $p_{\sigma}^2 = .003$).

**Stability of Facet Contributions.** As was the case with self-ratings, the ability of this study to provide an adequate assessment of the stability of facet contribution across organizations is limited due to the small number of organizations for which peer ratings were available. Thus, the findings presented in Table 17 should be interpreted with caution as data were only sampled from 10 organizations. Given this qualifying statement, the findings presented in Table 17 suggest that the stability of facet contributions was quite poor for many facets. For example, sampling error only accounted for 39.7% (Altruism) and 37.9% (Civic Virtue) of the raw variance in estimates of the individual facet’s relative contribution across organizations. Clearly these values fell well below Hunter and Schmidt’s (1990) 75% rule of thumb and, in both cases, $Q$-tests of homogeneity revealed significant differences between the raw and error variance estimates generated for the individual facet’s contribution across organizations (Altruism, $Q = 25.2$, $p <$
The individual facet accounted for between 13.1% and 64.9% of total variance in Altruism across organizations ($MP_{\sigma^2} = .448$, $ISDR p_{\sigma^2} = .375$), and between 0.1% and 38.8% of total variance in Civic Virtue across organizations ($MP_{\sigma^2} = .248$, $ISDR p_{\sigma^2} = .686$).

In addition to the individual facet, the group facet also exhibited instability in its relative contribution to total variance for both Altruism and Civic Virtue. Specifically, sampling error only accounted for 41.8% (Altruism) and 73.4% (Civic Virtue) of the raw variance in estimates of the group facet’s contribution across organizations. Although both of these values fell below Hunter and Schmidt’s (1990) 75% rule of thumb, $Q$-tests for homogeneity revealed that only in the case of Altruism was there a significant difference between the raw and error variance estimates generated for the group facet’s relative contribution across organizations (Altruism, $Q = 23.9$, $p < .01$; Civic Virtue, $Q = 13.6$, n.s.). The group facet accounted for between 2.1% and 48.1% of total variance in Altruism across organizations ($MP_{\sigma^2} = .166$, $ISDR p_{\sigma^2} = .316$), and between 1.0% and 26.1% of total variance in Civic Virtue across organizations ($MP_{\sigma^2} = .095$, $ISDR p_{\sigma^2} = .308$).

The stability of the remaining facets depended on the dimension of OCB that was considered. For Altruism, the relative contribution of the remaining facets (i.e., behavior, group-by-behavior, and individual-by-behavior/residual) all appeared to be quite stable, with sampling error accounting for nearly all of their raw variance in estimates across organizations. With regard to Civic Virtue however, both the behavior facet and individual-by-behavior/residual term exhibited instability in their relative contributions across organizations. Specifically, sampling error only accounted for 71.3% (behavior facet) and 40.9% (individual-by-behavior/residual) of the raw variance in estimates of these terms’ relative contributions to Civic Virtue across organizations. Although both of these values fell below Hunter and Schmidt’s (1990) 75% rule of
thumb, $Q$-tests for homogeneity revealed that only in the case of the individual-by-behavior/residual term was there a significant difference between the raw and error variance generated for the term’s relative contribution across organizations ($Q = 24.5, p < .01$). The individual-by-behavior/residual term accounted for between 24.8% and 79.7% of total variance in Civic Virtue across organizations ($MP_{\sigma^2} = .511, ISDR p_{\sigma^2} = .443$), whereas the behavior facet accounted for between 1.5% and 26.5% of total variance in Civic Virtue across organizations ($MP_{\sigma^2} = .08, ISDR p_{\sigma^2} = .362$).

**Supervisor Ratings**

Table 18 presents the results of the comparative fit analyses for architectures with and without the behavior facet.

<table>
<thead>
<tr>
<th>Dimension</th>
<th>$-2L\Delta$</th>
<th>AIC $\Delta$</th>
<th>$\sigma^2$ Residual $\Delta$</th>
<th>$p_{\sigma^2}$ Residual $\Delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Altruism</td>
<td>486.8†</td>
<td>-267.3</td>
<td>.070</td>
<td>.078</td>
</tr>
<tr>
<td>Civic Virtue</td>
<td>631.0†</td>
<td>-312.5</td>
<td>.190</td>
<td>.182</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>381.1†</td>
<td>-187.6</td>
<td>.110</td>
<td>.099</td>
</tr>
<tr>
<td>Courtesy</td>
<td>783.2†</td>
<td>-388.5</td>
<td>.130</td>
<td>.121</td>
</tr>
<tr>
<td>Sportsmanship</td>
<td>458.7†</td>
<td>-236.4</td>
<td>.150</td>
<td>.094</td>
</tr>
</tbody>
</table>

*Note.* $df = 3$ for all dimensions. All differences were calculated by taking values from architectures with behavior and subtracting them from values from architectures without behavior. † $p < .001$

As revealed by Table 18, the architectures including the behavior facet provided a significantly better fit to the data than architectures that excluded the behavior facet. Specifically, including the behavior facet in the architecture along with its corresponding interactions (group-by-behavior and organization-by-behavior) resulted in 22.0% (Altruism), 34.9% (Civic Virtue), 20.9% (Conscientiousness), 30.9% (Courtesy), and 26.6% (Sportsmanship) reductions in the relative contribution of residual variance to total variance in the five OCB dimensions, and 20.6%
(Altruism), 32.2% (Civic Virtue), 20.8% (Conscientiousness), 29.3% (Courtesy), and 27.6% (Sportsmanship) reductions in the absolute contribution of residual variance for these dimensions. As was the case with self and peer ratings, these findings indicate that architectures that ignore behavior-related variation would not suffice for describing the variance underlying dimensions of the OCB construct. Given these findings, Table 19 only presents parameter estimates for variance architectures where behavior was included as a facet.

Magnitude of Facet Contributions. The variance architectures underlying different dimensions of OCB as revealed by supervisor ratings shared a number of similarities and differences. For example, the largest source of variation for four of the five OCB dimensions was the individual facet. The individual facet accounted for between 26.3% (Civic Virtue) and 46.3% (Altruism) of the total variance in each dimension. Only in the case of Civic Virtue did the relative and absolute contribution of another facet exceed that of the individual facet. Specifically, the individual-by-behavior/residual term accounted for 33.9% of the variance in Civic Virtue. For four of the five OCB dimensions examined (the exception being Civic Virtue), the next largest source of variation was the individual-by-behavior/residual term. This term accounted for between 25.6% (Courtesy) and 37.4% (Conscientiousness) of the total variance in each dimension.

Another similarity that was evident among architectures revealed by supervisor ratings was that the group facet consistently emerged as a pertinent source of variation (similar to peer ratings). Specifically, the group facet accounted for between 8.5% (Conscientiousness) and 17.6% (Civic Virtue) of the total variance in each OCB dimension. In contrast to both peer ratings and self-ratings, the group-by-behavior term, and to a lesser extent organization facet, accounted for significant proportions of variation in each OCB dimension. Specifically, the group-by-
Table 19. Variance Architecture Summary for Supervisor Ratings

<table>
<thead>
<tr>
<th>Facet</th>
<th>Facet Contributions</th>
<th>Facet Stability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Absolute</td>
<td>Relative</td>
</tr>
<tr>
<td></td>
<td>$\sigma^2$</td>
<td>$\sigma^2$</td>
</tr>
<tr>
<td></td>
<td>90% L</td>
<td>90% U</td>
</tr>
<tr>
<td>---------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>Altruism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>i:g:o</td>
<td>.450</td>
<td>.422</td>
</tr>
<tr>
<td>g:o</td>
<td>.135</td>
<td>.106</td>
</tr>
<tr>
<td>o</td>
<td>.071</td>
<td>.044</td>
</tr>
<tr>
<td>b</td>
<td>.012</td>
<td>.003</td>
</tr>
<tr>
<td>g x b</td>
<td>.035</td>
<td>.031</td>
</tr>
<tr>
<td>o x b</td>
<td>.002</td>
<td>.001</td>
</tr>
<tr>
<td>i:g:o x b (res)</td>
<td>.267</td>
<td>.259</td>
</tr>
</tbody>
</table>

| Civic Virtue|        |        |        |        |
| i:g:o   | .310   | .282   | .342   | .263   | .233   | .292   | 29  | .273   | .196 | 100.0 |
| g:o     | .207   | .165   | .272   | .176   | .148   | .205   | 29  | .124   | .303 | 100.0 |
| o       | .040   | .019   | .251   | .034   | .019   | .056   | .     | .     | .     | .     |
| b       | .054   | .025   | .333   | .046   | .006   | .255   | 29  | .039   | .114 | 100.0 |
| g x b   | .142   | .122   | .169   | .120   | .094   | .148   | 29  | .118   | .233 | 100.0 |
| o x b   | .027   | .017   | .052   | .023   | .012   | .035   | .     | .     | .     | .     |
| i:g:o x b (res) | .400  | .382   | .418   | .339   | .315   | .363   | 29  | .377   | .236 | 100.0 |

| Conscientiousness|        |        |        |        |
| i:g:o   | .436   | .400   | .477   | .389   | .360   | .418   | 24  | .393   | .288 | 100.0 |
| g:o     | .095   | .065   | .159   | .085   | .065   | .107   | 24  | .084   | .197 | 100.0 |
| o       | .062   | .034   | .191   | .055   | .032   | .091   | .     | .     | .     | .     |
| b       | .013   | .005   | .118   | .012   | .001   | .071   | 24  | .022   | .085 | 100.0 |
| g x b   | .089   | .075   | .109   | .080   | .057   | .104   | 24  | .075   | .136 | 100.0 |
| o x b   | .007   | .003   | .043   | .006   | .000   | .015   | .     | .     | .     | .     |
| i:g:o x b (res) | .419  | .401   | .436   | .374   | .350   | .397   | 24  | .395   | .234 | 100.0 |
Table 19 - Continued

<table>
<thead>
<tr>
<th>Facet</th>
<th>Facet Contributions</th>
<th>Facet Stability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Absolute</td>
<td>Relative</td>
</tr>
<tr>
<td></td>
<td>( \sigma^2 )</td>
<td>( \sigma^2 ) (_{90% L} )</td>
</tr>
<tr>
<td>Courtesy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>i:g:o</td>
<td>.471 .434 .513</td>
<td>.444 .415 .472</td>
</tr>
<tr>
<td>g:o</td>
<td>.137 .101 .207</td>
<td>.129 .104 .157</td>
</tr>
<tr>
<td>o</td>
<td>.037 .010 9.460</td>
<td>.035 .018 .063</td>
</tr>
<tr>
<td>b</td>
<td>.015 .007 .396</td>
<td>.014 .002 .063</td>
</tr>
<tr>
<td>g x b</td>
<td>.082 .071 .098</td>
<td>.078 .056 .101</td>
</tr>
<tr>
<td>o x b</td>
<td>.033 .023 .058</td>
<td>.031 .019 .046</td>
</tr>
<tr>
<td>i:g:o x b (res)</td>
<td>.286 .275 .298</td>
<td>.270 .245 .294</td>
</tr>
<tr>
<td>Sportsmanship</td>
<td></td>
<td></td>
</tr>
<tr>
<td>i:g:o</td>
<td>.640 .587 .701</td>
<td>.391 .359 .422</td>
</tr>
<tr>
<td>g:o</td>
<td>.170 .119 .268</td>
<td>.104 .080 .131</td>
</tr>
<tr>
<td>o</td>
<td>.248 .147 .559</td>
<td>.151 .088 .245</td>
</tr>
<tr>
<td>b</td>
<td>.006 .003 2.511</td>
<td>.003 .000 .023</td>
</tr>
<tr>
<td>g x b</td>
<td>.142 .122 .170</td>
<td>.087 .062 .113</td>
</tr>
<tr>
<td>o x b</td>
<td>.009 .004 .097</td>
<td>.005 .000 .014</td>
</tr>
<tr>
<td>i:g:o x b (res)</td>
<td>.423 .405 .442</td>
<td>.259 .231 .285</td>
</tr>
</tbody>
</table>

Note. \(^{a}\) Value falls below Hunter-Schmidt’s 75% cut-off. \(^{b}\) Observed and error variance significantly different based on \(Q\)-test of homogeneity.

behavior facet accounted for between 3.6% (Altruism) and 12.0% (Civic Virtue) of the total variance in dimensions, whereas the organization facet accounted for between 3.4% (Civic Virtue) and 15.1% (Sportsmanship) of the total variance in dimensions. Neither the behavior facet, nor organization-by-behavior facet made any substantial contribution to the total variance in any dimension examined.

**Stability of Facet Contributions.** Depending on the dimension being examined, supervisor ratings were available for between 19 and 39 organizations. The findings presented in Table 19 reveal that facet contributions across organizations were relatively stable. For example, all raw
variation (across organizations) in the relative contribution of all facets in the architectures underlying Civic Virtue and Conscientiousness was accounted for by sampling error. For Altruism, Courtesy, and Sportsmanship, three of the five facets in their architectures (behavior, group-by-behavior, and individual-by-behavior/residual) exhibited similarly high levels of stability. However, for these latter three dimensions both the individual facet and, to a lesser extent, the group facet exhibited signs of instability in their relative contributions to total variance across organizations.

For example, the individual facet accounted for between (a) 19.7% and 72.3% of total variance in Altruism across organizations ($M_{p^2} = .493, ISDR_{p^2} = .314$), (b) 6.3% and 81.1% of total variance in Courtesy across organizations ($M_{p^2} = .453, ISDR_{p^2} = .355$), (c) 18.2% and 69.2% of total variance in Sportsmanship across organizations ($M_{p^2} = .459, ISDR_{p^2} = .313$). Sampling error only accounted for 58.9% (Altruism), 53.7% (Courtesy), and 56.2% (Sportsmanship) of the raw variance in estimates of the individual facet’s relative contribution across organizations. Clearly these values below Hunter and Schmidt’s (1990) 75% rule of thumb, and in all three cases $Q$-tests of homogeneity revealed significant differences between the raw and error variance estimates generated for the individual facet’s contribution across organizations (Altruism, $Q = 66.2, p < .01$; Courtesy, $Q = 37.3, p < .01$; Sportsmanship, $Q = 32.0, p < .05$).

Although the group facet also exhibited signs of instability in its relative contribution to total variance across organizations, such instability appeared to be greater for Sportsmanship relative to both Altruism and Courtesy. Specifically, sampling error accounted for 73.9% (Sportsmanship), 81.6% (Altruism), and 81.8% (Courtesy) of the raw variance in estimates of the group facet’s relative contribution across organizations. Only the value for Sportsmanship falls below Hunter and Schmidt’s (1990) 75% rule of thumb. However, a $Q$-test for homogeneity
revealed no significant difference between the raw and error variance estimates generated for the group facet’s contribution to Sportsmanship across organizations ($Q = 24.4, \text{n.s.}$). The group facet accounted for between 0.1% and 52.6% of total variance in Sportsmanship across organizations ($Mp_{\sigma^2} = .103, ISDR p_{\sigma^2} = .270$).

Addressing Architecture-Level and Facet-Level Research Questions

The sections above provided details on architectures underlying each dimension of OCB as revealed by different rating methods. In the present section, the focus is primarily on comparison of architectures at both the facet- and architecture-levels, taking into consideration that architectures may be compared across dimensions, across rating methods, or across combinations of both (i.e., comparing architectures underlying specific dimension-method combinations). In the process of making these comparisons, many of the OCB-specific research questions raised in the introduction (see Table 5) are directly addressed.

Architecture-Level Comparisons- Relative Facet Contributions

Table 20 presents $ASI-RC$ values corresponding to each pairwise comparison of all “i:g:o x b” architectures examined in this study. As alluded to earlier, this table is very similar in structure to a traditional MTMM matrix used in the context of construct validation (Campbell & Fiske, 1959). Interpreting an $ASI-RC$-based MTMM matrix is similar to interpreting a correlation-based MTMM matrix in terms of the ideal pattern of values. In the $ASI-RC$-based MTMM matrix, it would generally be desirable to find high monodimension-heteromethod $ASI-RC$ values relative to both (a) heterodimension-monomethod $ASI-RC$ values, and (b) heterodimension-heteromethod $ASI-RC$ values. Such a pattern would suggest that there exists a common architecture underlying each dimension of OCB that does not vary as a function of the method of measurement used to reveal it. Conversely, it would also generally be desirable to find low heterodimension-
<table>
<thead>
<tr>
<th>Method</th>
<th>Dim.</th>
<th>Self-Ratings</th>
<th></th>
<th>Peer Ratings</th>
<th></th>
<th>Supervisor Ratings</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Alt.</td>
<td>C.V.</td>
<td>Court.</td>
<td>Alt.</td>
<td>C.V.</td>
<td>Court.</td>
</tr>
<tr>
<td>Self-Ratings</td>
<td></td>
<td>Alt.</td>
<td>.950 (.014)</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td></td>
<td>C.V.</td>
<td>.854 (.042)</td>
<td>.848 (.043)</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>Peer Ratings</td>
<td></td>
<td>Alt.</td>
<td>.668 (.095)</td>
<td>.632 (.105)</td>
<td>.557 (.127)</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td></td>
<td>C.V.</td>
<td>.809 (.055)</td>
<td>.811 (.054)</td>
<td>.766 (.067)</td>
<td>.672 (.094)</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>Supervisor Ratings</td>
<td></td>
<td>Alt.</td>
<td>.661 (.097)</td>
<td>.625 (.107)</td>
<td>.541 (.131)</td>
<td>.916 (.024)</td>
<td>.676 (.093)</td>
</tr>
<tr>
<td></td>
<td>C.V.</td>
<td>.685 (.090)</td>
<td>.670 (.094)</td>
<td>.626 (.107)</td>
<td>.746 (.073)</td>
<td>.790 (.060)</td>
<td>.761 (.068)</td>
</tr>
<tr>
<td></td>
<td>Con.</td>
<td>.763 (.068)</td>
<td>.727 (.078)</td>
<td>.644 (.102)</td>
<td>.806 (.055)</td>
<td>.781 (.063)</td>
<td>.853 (.042)</td>
</tr>
<tr>
<td></td>
<td>Court.</td>
<td>.686 (.090)</td>
<td>.650 (.100)</td>
<td>.565 (.124)</td>
<td>.882 (.034)</td>
<td>.689 (.089)</td>
<td>.928 (.021)</td>
</tr>
<tr>
<td></td>
<td>Sports.</td>
<td>.639 (.103)</td>
<td>.603 (.113)</td>
<td>.528 (.135)</td>
<td>.793 (.059)</td>
<td>.664 (.096)</td>
<td>.867 (.038)</td>
</tr>
</tbody>
</table>

*Note.* Values outside of the parentheses are the ASI-RC values for the two architectures being compared. Values inside the parentheses are the mean absolute differences between the relative contributions of the common facets of the architectures being compared. Recall ASI-RC values range from 0 to 1, where higher values indicate a higher degree of similarity in the dispersion of variance across the facets of the two architectures being compared. The “Dim” column indicates the OCB dimensions being examined. Monodimension-heteromethod ASI-RC values and their corresponding mean absolute differences are italicized.
monomethod *ASI-RC* values relative to both (a) monodimension-heteromethod *ASI-RC* values and (b) heterodimension-heteromethod *ASI-RC* values. Such a pattern would suggest that different dimensions of OCB have distinct architectures underlying them and that method variance does not overwhelm the distinctiveness of these architectures.

In the present study, little evidence emerged for the notion that the different dimensions of OCB have distinct architectures. For example, the average monodimension-heteromethod *ASI-RC* values for Altruism (*M* = .748; *Range* = .661 to .916) and Civic Virtue (*M* = .757; *Range* = .670 to .811) were lower than the average heterodimension-monomethod *ASI-RC* values for both self-ratings (*M* = .884, *Range* = .848 to .950) and supervisor ratings (*M* = .842, *Range* = .755 to .928). Nevertheless, the average Civic Virtue-heteromethod *ASI-RC* value was higher than the average heterodimension-heteromethod *ASI-RC* values for all method pairings [self-peer (.691), self-supervisor (.643), peer-supervisor (.755)], whereas the average Altruism-heteromethod *ASI-RC* value was higher than two of the three average heterodimension-heteromethod *ASI-RC* values (self-peer, self-supervisor).

Although these patterns of *ASI-RC* values suggest little evidence of the distinctness of the architectures underlying the different dimensions of OCB (at least in terms of the relative dispersion of variance across their facets), they do suggest strong method effects. Specifically, the average heterodimension-monomethod *ASI-RC* values were higher than both (a) the average monodimension-heteromethod *ASI-RC* values and (b) the average heterodimension-heteromethod *ASI-RC* values. Thus, any differences that exist between dimensions of OCB at the architecture-level appear to be greatly overshadowed by method effects.

In addition to addressing the last two OCB research questions posed in Table 5, other architecture-level questions can also be addressed with results presented in Table 20. For example, recall the research question that regarded whether architectures of OCB revealed by
peer and supervisor ratings would generally be more similar to each other than to architectures of OCB revealed by self-ratings. Based on the current results, it appears that the answer to this question depends on the dimension of OCB being considered. For example, although the average heterodimension-heteromethod *ASI-RC* value for peer-supervisor pairing was higher than the corresponding average *ASI-RC* values for both supervisor-self and peer-self pairings, this was not the case for all dimensions. Specifically, although the Altruism-heteromethod *ASI-RC* value for the peer-supervisor pairing (.916) was indeed higher than the corresponding *ASI-RC* values for both the supervisor-self (.661), and peer-self (.668) pairing, this was not true for Civic Virtue. The Civic Virtue-heteromethod *ASI-RC* value for the peer-supervisor pairing (.790) was higher than the corresponding *ASI-RC* value for the supervisor-self paring (.670), but not the value for the peer-self pairing (.811).

Another research question raised earlier regarded the possibility that the architectures underlying OCB dimensions nested within each higher-order OCB dimension (i.e., OCBI and OCBO) would be more similar to each other than the architectures of the OCB dimensions in the other higher-order OCB dimension. Specifically, based on the OCBI/OCBO distinction, architectures underlying Altruism and Courtesy should be more similar to each other than to architectures underlying Civic Virtue, Conscientiousness, or Sportsmanship. Conversely, the latter three dimensions’ architectures should be more similar to each other than to those underlying Altruism or Courtesy. Given the strong method effects revealed by the MTMM interpretation of results presented above, as well as the lack of data on all dimensions of the OCB construct for self and peer ratings, answering this question with the current data is difficult. To simplify the answer to this question comparisons are discussed within each rating method, to eliminate the “noise” created by method effects.
Generally, the results were mixed, but positive with regard to the expected pattern of similarities. For self-ratings, the Altruism-Courtesy ASI-RC value (.854) was only slightly higher than the Civic Virtue-Courtesy value (.848) and actually lower than the Altruism-Civic Virtue ASI-RC value (.950). For supervisor ratings, the findings were more positive. Specifically, the Altruism-Courtesy ASI-RC value (.928) was higher than all other Altruism-OCBO dimension ASI-RC values \( (M = .827; \text{Range} = .761 \text{ to } .867) \), as well as higher than all other Courtesy-OCBO dimension ASI-RC values \( (M = .853; \text{Range} = .809 \text{ to } .875) \). The Civic Virtue-Conscientiousness ASI-RC value (.818) was higher than the average Civic Virtue-OCBI dimension ASI-RC value \( (M = .785; \text{Range} = .761 \text{ to } .809) \), but lower than the average Conscientiousness-OCBI dimension ASI-RC value \( (M = .864; \text{Range} = .853 \text{ to } .874) \). The Conscientiousness-Sportsmanship ASI-RC value (.876) was higher than both the average Conscientiousness-OCBI dimension ASI-RC value (.864) and the average Sportsmanship-OCBI dimension ASI-RC value (.871). On the other hand, the Civic Virtue-Sportsmanship ASI-RC value (.755) was lower than both the average Civic Virtue-OCBI dimension ASI-RC value (.785) and the average Sportsmanship-OCBI dimension ASI-RC value \( (M = .871; \text{Range} = .867 \text{ to } .875) \).

**Architecture-Level Comparisons- Stability of Relative Facet Contributions**

In addition to making architecture-level comparisons based on the similarity in the profiles of relative facet contributions (as indexed by ASI-RC values), one may also compare the similarity in profiles of “true” variance around such facet contributions across organizational contexts using ASI-RS values (comparison type IV-A in Table 4). Table 21 presents the ASI-RS values corresponding to each pairwise comparison of all “i:g:o x b” architectures examined in this study.

As detailed in Table 21, the stability of the architectures underlying different dimensions of OCB as well as those architectures revealed by each rating method appeared to be very similar.
Table 21. An ASI-RS-based MTMM Matrix

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-Ratings</td>
<td>Alt.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td></td>
<td>C.V.</td>
<td>1.000</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td></td>
<td>Court.</td>
<td>.993</td>
<td>.993</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>Peer Ratings</td>
<td>Alt.</td>
<td>.951</td>
<td>.951</td>
<td>.944</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td></td>
<td>C.V.</td>
<td>.897</td>
<td>.897</td>
<td>.905</td>
<td>.920</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>Supervisor Ratings</td>
<td>Alt.</td>
<td>.981</td>
<td>.981</td>
<td>.973</td>
<td>.970</td>
<td>.917</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td></td>
<td>C.V.</td>
<td>1.000</td>
<td>1.000</td>
<td>.993</td>
<td>.951</td>
<td>.897</td>
<td>.981</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td></td>
<td>Con.</td>
<td>1.000</td>
<td>1.000</td>
<td>.993</td>
<td>.951</td>
<td>.897</td>
<td>.981</td>
<td>1.000</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td></td>
<td>Court.</td>
<td>.977</td>
<td>.977</td>
<td>.970</td>
<td>.974</td>
<td>.920</td>
<td>.995</td>
<td>.977</td>
<td>.977</td>
<td>.977</td>
<td>.993</td>
</tr>
<tr>
<td></td>
<td>Sports.</td>
<td>.977</td>
<td>.977</td>
<td>.969</td>
<td>.974</td>
<td>.921</td>
<td>.996</td>
<td>.977</td>
<td>.977</td>
<td>.977</td>
<td>.993</td>
</tr>
</tbody>
</table>

*Note.* Recall ASI-RS values range from 0 to 1, where higher values indicate a higher degree of similarity in the stability of the dispersion of variance across the facets of the two architectures being compared. The “Dim” column indicates the OCB dimensions being examined. Monodimension-heteromethod ASI-RS values are italicized.
For example, the average ASI-RS value for all dimension pairings based on self-ratings was .995 (Range = .993 to 1.0), for peer ratings this value was .920, and for supervisor ratings the average ASI-RS value was .985 (Range = .977 to 1.0). Furthermore, the ASI-RS values were also very high for comparisons of each OCB dimension with the other dimensions across all possible method pairings. Specifically, the average ASI-RS values associated with Altruism ($M = .968; \text{Range} = .897 \text{ to } 1.0$), Civic Virtue ($M = .955; \text{Range} = .897 \text{ to } 1.0$), Conscientiousness ($M = .975; \text{Range} = .897 \text{ to } 1.0$), Courtesy ($M = .967; \text{Range} = .905 \text{ to } .995$), and Sportsmanship ($M = .973; \text{Range} = .921 \text{ to } .996$) were all very high. Given the high degree of similarity exhibited among architectures with regard to the stability of their relative facet contributions, providing answers to the architecture-level OCB research questions based on ASI-RS values does not appear worthwhile (i.e., due to uniformly high levels of similarity). Specifically, because all of the the ASI-RS values are so high, there appears to be very little utility in distinguishing between the architectures in terms of the ASI-RS index in the present study.

The great deal of similarity in all ASI-RS values reveals a potential weakness of comparing architectures using the ASI-RS index. Specifically, the index may not be particularly sensitive for making architecture-level comparisons for situations in which one does not have large samples of units of each facet within each organization. Small samples of units for a facet within an organization will inflate estimates of variance in facet contributions across organizations that is due to sampling error. Thus, to the extent that only small number of units are available, it may appear that no true variance in facet contributions across organizations exists. Although this is a potential problem when examining stability at both the facet- and architecture-levels, the problem is arguably compounded at the architecture-level. Specifically, if one has small samples of units of one facet within each organization, it may necessitate that samples of units of other facets within each organization are also small (e.g., small numbers of individuals
necessitate small number of groups, jobs, etc.). In calculating the ASI-RS index this would translate into taking more differences between pairs of estimated true variability in facet contributions that are near 0.

Facet-Level Comparisons- Absolute Facet Contributions

Figure 4 presents a graph of absolute facet contributions and their 90% confidence intervals for those OCB dimensions that were assessed by two or more rating methods. Although the information in this figure is presented in Tables 15, 17, and 19, it provides a graphical representation of those results and facilitates comparison of absolute facet contributions across both dimension and rating method (i.e., the focus of comparisons in this section).

Recall that several OCB research questions summarized in Table 5 were geared toward facet-level comparisons. For example, one such facet-level research question regarded whether the individual, group, and organization facets would account for less variance in self-ratings compared to both peer and supervisor ratings. Based on present findings there appears be good support for this claim. Specifically, with the exception of the individual-level facet for Civic Virtue, the individual, group, and organization facets all have smaller variance components for self-ratings compared to both peer and supervisor ratings. These findings, along with the means presented in Tables 9, 10, 11, are consistent with the notion that self-raters may be inflating their ratings. Such inflation in turn leads to decreased variance at the individual and aggregate (group and organization) levels due to a floor effect (recall the lower bound of the rating scale is 1, however, if inflation occurs than the lower observed bound is actually greater than 1).

As follow-up to the research question addressed in the previous paragraph, an additional question was posed regarding whether differences in contributions of the aforementioned facets,
Figure 4. Absolute Facet Contributions as Revealed by Different Rating Methods

and in particular the individual facet, would be more salient in comparisons of self and supervisor ratings. Based on the results presented in Figure 4, no support for this claim was provided.

Specifically, whereas the variance component of the individual-facet for self-ratings of Altruism ($\sigma_{i,g,o}^2 = .309$) was significantly smaller than the variance component of the individual facet for both peer ratings ($\sigma_{i,g,o}^2 = .670$), and supervisor ratings ($\sigma_{i,g,o}^2 = .450$), the distance between the self and peer values was greater than the distance between the self and supervisor values. With regard to Civic Virtue, the variance component of the individual facet for self-ratings ($\sigma_{i,g,o}^2 = .446$), was actually greater than the individual-facets variance component for both supervisor ratings ($\sigma_{i,g,o}^2 = .310$) and peer ratings ($\sigma_{i,g,o}^2 = .274$).

Another facet-level research question posed in Table 5 regarded whether the individual-by-behavior/residual term and indeed behavioral sources of variation in general account for more variance among self-ratings of OCB compared to both peer and supervisor ratings. The logic behind this question was that (a) self-raters have a reduced cognitive load on the ratings task relative to peers and supervisors, who often must rate more than just, and thus may more finely distinguish between individual instances of their behavior themselves (Lord & Maher, 1991), and (b) self-raters have more motive to do so (inflating their rating on behaviors they believe make them appear most effective; Murphy and Cleveland, 1995). For the most part, the findings of this study were consistent with this logic. Specifically, the individual-by-behavior/residual and organization-by-behavior terms accounted for significantly more variance among self-ratings relative to both peer and supervisor ratings. However, this pattern (i.e. self-ratings’ variance components being greater than peer and supervisor ratings’ variance components) did not hold for either the behavior facet, or the group-by-behavior term. As mentioned earlier, it could be that variance due to the behavior facet among the self-ratings was effectively cancelled out if different OCBs were differentially valued among the different organizations examined. Consistent with
this reasoning, when separate architectures for each organization based on self-ratings, the behavior-facet accounted for substantially more variance than it did among both peer ratings and supervisor ratings.

A follow-up question was also posed that regarded the magnitude of the differences in the individual-by-behavior/residual term between the self and supervisor architectures compared to the self and peer architectures. Differences on this facet were expected to be greater when comparing self and supervisor architectures than when comparing self and peer architectures because the cognitive load on supervisors is typically greater than that for peers on the rating task. This increased cognitive load, may lead supervisors to rely on heuristics and general impressions of individual performance, thus potentially leading them to make grosser distinctions between individuals engagement in specific behaviors relative to either peer or self-raters. Consistent with this reasoning, greater differences were found among variance components for the individual-by-behavior facet underlying the architectures of Altruism and Civic as revealed by self-ratings ($\sigma_{\text{Altruism}}^2 = .472; \sigma_{\text{Civic Virtue}}^2 = .691$) and supervisor ratings ($\sigma_{\text{Altruism}}^2 = .267; \sigma_{\text{Civic Virtue}}^2 = .400$), compared to those revealed by self-ratings and peer ratings ($\sigma_{\text{Altruism}}^2 = .379; \sigma_{\text{Civic Virtue}}^2 = .546$). Taken together, these findings suggest that self-raters were making finer distinctions between specific instances of OCB compared to peer raters and even more so compared supervisors.

Facet-Level Comparisons- Stability of Relative Facet Contributions

The stability of facet contributions across organizations appeared to vary slightly across rating method, although for both self and peer ratings, only a small number of organizations were available to generate stability estimates, thus results should be interpreted with caution. For the most part, facet contributions based on self-ratings appeared quite stable. Not only were the inter-standard deviation range values around the average proportions of variance accounted for by facets across organizations generally low (Mean ISDR $p_\pi^2 = .137$, across facets and dimensions),
but nearly all of this variance could be explained by sampling error. Specifically, the average percentage of raw variance in facet contributions for self-ratings across organization that could be attributed to sampling error was 98.2%. These findings are consistent with suggestions raised earlier that there are some universal biases that may be exhibited by self-raters (e.g., impression management, self-enhancement) that may lead to more stable patterns of variation in the facets that contribute to construct variance as revealed by self-ratings.

Unlike the self-ratings, the contribution of the facets to the total variance in OCB based on peer ratings appeared to be much less stable, particularly for the Civic Virtue. First of all the inter-standard deviation range values tended to be higher for peer ratings ($\text{Mean ISDR p}^2 = 289$, across facets and dimensions) than self-ratings and much less of this variability appeared to be accounted for by sampling error. Specifically, the average percentage of raw variance in facet contributions across organizations for peer ratings that could be attributed to sampling error was 70.5%. These findings as well as the fact that Hunter and Schmidt’s (1990) 75% rule was violated for six of the ten facets examined among peer ratings of Altruism and Civic Virtue suggest that there is less stability in the contribution of the facets for peer ratings compared to self-ratings.

As for the stability facet contributions across organizations in architectures revealed by supervisor ratings, the facet contributions underlying Civic Virtue and Conscientiousness dimensions appeared quite stable (i.e., all of their observed variance was accounted for by sampling error). As for Altruism, Courtesy, and Sportsmanship, only the individual and group facets appeared to be unstable, as the variation in all other facets was accounted for by sampling error. Like self-ratings, the average inter-standard deviation range values tended to be lower than that of peer ratings ($\text{Mean ISDR p}^2 = .199$, across facets and dimensions). Additionally, more of this variability was accounted for by sampling error compared to peer ratings, as the average
percentage of raw variance in facet contributions for supervisor ratings across organizations that could be attributed to sampling error was 92.2%.

Discussion

Examining OCB through the lens of the variance architecture approach not only provided insight into the OCB construct and methods by which it is assessed, but also the utility and potential limitations of the approach in general. In the sections below, I discuss the insight this study has provided into each of these areas, the potential directions for future OCB research, and the use of the variance architecture approach as a research tool.

Insights for OCB Research

Data Gathering, Coding, and Measurement Designs

This study revealed that the available OCB data appear to be lacking in their ability to provide a full delineation of the OCB construct’s variance architecture. Specifically, careful examination of the literature and subsequent data collection efforts revealed that very little OCB data was available with regard to the culture, job, and occasion facets. Part of this was due to a lack of sampling across units of these facets (e.g., culture, and occasion), whereas some of this stemmed from the lack of coding on the part of researchers in managing their data sets (e.g., job-related information). Given that OCB has primarily been studied as an individual-level construct (and thus coding for other facets may have not been of much theoretical interest to researchers when conducting their studies), the lack of coding for many VAA facets outlined in this manuscript is somewhat understandable. Nevertheless, fully delineating a construct’s architecture has many benefits, and thus coding for these facets when they are tapped is an important step for future OCB researchers to take.

Related to this lack of facet information, is a lack of variation in terms of how the facets that were available for this study (i.e., organizations, groups, individuals, and behaviors) were
oriented to each other (e.g., nested or crossed). Given that many of the facets examined in this study were nested within one another, it was impossible to uniquely estimate the effects stemming from some of these facets. Uniquely estimating the contribution of each facet and their interactions with each other would require that the facets be fully crossed rather than nested. Although it is unlikely that fully crossed facets will ever be the norm (due to the fact that organizations are typically hierarchically structured entities), having a small body of such OCB studies would help generate estimates for the confounded sources of variation in the present study. Thus, future OCB researchers may wish to seek out contexts for research that allow for more crossing of VAA facets than was currently possible with the available OCB data.

Along similar lines, given how OCB ratings are usually obtained (i.e., self, peer, supervisor ratings), some facet effects could not be separated from effects arising from observers’ individual rating idiosyncrasies (e.g., group effects and observer effects for architectures revealed by supervisor ratings). To be able to uniquely interpret the variance stemming from such facets, measurement designs could be implemented that allow for the unique estimation of the facets that were confounded in the present study. For example, if multiple supervisors provided OCB ratings for group members or supervisors provided ratings for multiple groups, one could eliminate the confounding of group and observer facets that was inherent in supervisor ratings in the present study. Given the dominance of similar measurement designs underlying the collection of OCB data in the OCB literature, such alternative designs are not likely to be the norm. However, having a small body of OCB studies that rely on alternative measurement designs could begin to provide indicators of the amount of variance that is attributable to each of the facets that are confounded with the observer facet in this study (Lance, 1994; Scullen et al., 2000).
Based on the data that was made available for this study, the majority of variance accounted for in the OCB construct across dimensions appears to stem from the individual facet and individual-by-behavior/residual term. The strong showing of the individual facet to some extent confirms the apparent bias among OCB researchers of not coding for other facets. Indeed, such findings are consistent with classical test theory, in that these sources of variation are often viewed as the two components of observed score variance [i.e., true variance (variation across individuals) and error variance (variation from an interaction between individuals and scale items)] (Allen and Yen, 1978). Nevertheless, the strong findings with regard to these facets should be interpreted with some caution because variation across them may also be a function of facets that were unmeasured in the models examined (e.g., job, occasion). Furthermore, the variance attributable to other facets in OCB architectures was not trivial. Specifically, the average percentages of variance accounted for by the other facets examined, across all dimensions and rating method were 14.9% (self-ratings), 22.5% (peer ratings), and 30.7% (supervisor ratings).

Patterns of variation in facet contributions tended to differ most saliently by rating method (i.e., self, peer, supervisor), as opposed to OCB dimensions, and these method effects generally corresponded to patterns of variation that might be expected based on past research in the performance appraisal literature. For example, as expected, the individual, group, and organizational facets tended to account for smaller amounts of variance among self-ratings, relative to both peer and supervisor ratings. This finding, as well as the finding of higher means on the dimensions of OCB among the self-ratings are consistent with arguments for the existence of a self-enhancing bias among self-raters, which leads to a floor effect (i.e., the lower bound on self-ratings is higher than the lower bound of the scale due to a tendency to self-enhance), and
effectively reduces individual and aggregate-level variance in self-ratings relative to both peer and supervisor ratings.

Furthermore, the individual-by-behavior/residual term accounted for a greater amount of variance among self-ratings than among both peer and supervisor ratings. These findings are consistent with the notion that self-raters may be more likely to distinguish between specific instances of their own behavior than other raters who, due to their increased cognitive load on the rating task (particularly among supervisors), may use heuristics and general impressions to influence their ratings, and thus fail to distinguish between different instances of individuals behavior.

Although these findings are consistent with past explanations of the strengths and weaknesses of different sources of performance ratings offered by the performance appraisal literature, it is important to remember that this study does not provide a test of these explanations because no analysis of the primary causal agent occurred. With that said, future researchers in the OCB literature can provide further insight into this matter by conducting more rigorously controlled investigations to test whether or not the causal agents identified by the performances appraisal literature suggested to be at the root of these findings (e.g., self-raters impression management and inflation tendencies, supervisors’ use of heuristics and general impressions to reduce cognitive load) actually explain the differences in variance patterns revealed in the present study (Cook & Campbell, 1979).

In addition to examining how variance architectures underlying the OCB construct vary as a function of rating method, potential differences between dimensions of the OCB construct were also explored in the present study. Unlike arguments provided for why architectures revealed by different methods of assessing OCBs might differ, there was little a-priori expectation as to how the architectures revealed by different dimensions of OCB would differ. The only a-
priori expectation regarding similarities or differences of the architectures underlying OCB dimensions was that the architectures underlying OCB dimensions nested within each higher-order OCB dimension (i.e., OCBI and OCBO) would be more similar to each other than the architectures of the OCB dimensions in the other higher-order OCB dimension.

The results of this study with regard to this expectation were mixed for self-ratings, but more positive for supervisor ratings. Specifically, based on supervisor ratings, the architectures of Altruism and Courtesy were more similar to each other than they were to each of the other dimensions. Furthermore, architectures of Conscientiousness and Sportsmanship (as revealed by supervisor ratings) were more similar to each other than they were to architectures of Altruism and Courtesy. Unfortunately, there was not enough peer-ratings data to evaluate this expectation, because peer data was only available on the Altruism and Civic Virtue dimensions of OCB.

Expanding on the method-based and dimension-based comparisons made above, this study employed a multitrait-multimethod lens to elaborate the similarities and differences of architectures underlying OCB dimensions as revealed by different rating methods. At the architecture-level these multidimension-multimethod comparisons provided little evidence of either differences in patterns of dispersions of variance across facets or the relative stability of such dispersion across organizations as a function of the OCB dimension examined. As alluded to above, these architecture-level comparisons suggested that rating method effects overwhelmed any differences that may have been observed in the architectures of different dimensions of OCB. Essentially, what this means is that establishing a common underlying architecture for a given OCB dimension that reveals the true dispersion of variance across facets will require reconciling the differences observed between the architectures revealed by different methods. Future OCB research can be conducted to test specific causes of the differences observed among architectures underlying OCB revealed by different rating method.
Future Implementation of the Variance Architecture Approach

In the process of conducting the present investigation, several difficulties were experienced in trying to delineate a detailed architecture of the OCB construct that may speak to future efforts at delineating the architectures underlying other constructs (e.g., job satisfaction, work values, etc.), as well as the success of the VAA in general. These difficulties revolve around three issues: (a) availability of data for VAA types of analyses, (b) computational issues regarding the estimation of VA parameters, and (c) refining statistical methods for making comparisons among variance architectures.

Availability of VAA-Related Data

Obtaining data for use in the present study was quite challenging. Because no standards for data sharing and data coding are universally in place among organizational researchers, obtaining data, and moreover data that is sufficiently coded, is a major difficulty in applying this approach. Adding to this difficulty is the breadth of data that is needed to delineate a construct’s architecture. One way to minimize the impact of such difficulties is to view the process of delineating a constructs’ variance architecture as an on-going, iterative process, rather than simply a set of estimated parameters based on a one time sampling of the research literature surrounding a construct.

As evidenced by the present study’s investigation of the OCB construct, there simply was not enough data acquired from the current OCB literature to fully delineate architectures underlying OCB. Such a finding suggests, that full delineation of the OCB constructs architecture (or any construct for that matter) would require an on-going process that allows for further refinement and detailing of the variance architectures underlying the construct. Ideally, the foundation of such an iterative-VAA research process would be a centralized database of VAA-related data surrounding a given construct, that researchers could contribute data to on an on-
going basis as they proceed through their individual programs of research. Eventually, enough data would be contributed by researchers to provide a very rich sampling of the facets of a construct’s variance architecture.

Maintaining such a database may help alleviate (over time) many of the problems that arose with the current investigation of OCB and would likely arise in the investigation of other constructs as well. For example, in addition to the data collection and coding problem alluded to earlier, another problem that was revealed in attempting to delineate the architectures in the present study regarded the confounding of facet effects. This confounding arose from both the typical measurement designs used to gather the data (e.g., groups and observers for supervisor ratings) or typical structuring of organizations examined (e.g., hierarchically nested facets).

Having a continual outlet for VAA-related information such as a centralized database will allow researchers to submit data sets that can uniquely address such problems. For example, studies where facets that are traditionally nested (e.g., individuals and groups) are crossed and where alternative measurement designs are used that allow for the unique estimation of facets (e.g., multiple observers rating each individual) can help alleviate such problems. Although few of these studies may be available at any single time, over time as the literature surrounding a construct becomes broader and varied, the likelihood of finding such studies and including them in VAA-types of analyses will arguably increase.

In addition to the benefits of establishing a centralized database for VAA-related construct information, viewing the VAA as an iterative process is intuitive. For example, it is easy to conceptualize the linear VAA-research process described above in terms of a repeating cycle. First, with a constantly maintained database, the data-gathering phase of the VAA research process would be on-going and constant. Researchers could contribute data to such a centralized database any time they wish. The design specification phase (i.e., laying out architectural designs
based on the data currently available) would be undertaken whenever substantial amounts of construct data were added to the database for a given construct. Lastly, the data analysis phase would delineate architectures for the designs laid out in the previous phase, to produce continual refinements to a construct’s architecture on an annual basis.

The refinement of a construct's architecture would occur in two important ways through this iterative research process. First, refinement would occur in the degree to which one can generate stable estimates of the contribution of each facet in the architecture. Specifically, the more data one obtains with regard to a facet, the lower the standard errors of the estimate of that facet's contribution will become over time. Second, refinement would also occur in that it would be likely a wider variety of data will become available over time that will increasingly allow for the estimation of the contributions of more facets and their interactions. Specifically, although data on a limited number of facets may be available for initial delineation of a construct's architecture (e.g., facets in the present study), after several years of continued research it is likely that more facets will become available. The result of this added data will allow for a more detailed description of a construct's variance architecture.

*Computational Issues in the VAA*

One of the largest problems associated with carrying out the present study was the amount of computing time and power required to generate REML-based estimates for the variance components underlying each of the architectures examined. Table 22 presents computation times for the primary analyses conducted in the present study.

The times taken to run architectures based on supervisor ratings was substantial and this was even after reducing the overall number of cases to smaller numbers than were originally obtained. Indeed, many supervisor-based architectures could not be run without cutting the sample size down. These findings are particularly disturbing given the variance architecture
Table 22. Computation Times for Architecture Analyses in the Present Study (hrs:min:sec)

<table>
<thead>
<tr>
<th>Source</th>
<th>Target</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>i:g:o</td>
</tr>
<tr>
<td>Self-Ratings</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>All</td>
<td>0:01:17</td>
</tr>
<tr>
<td></td>
<td>Altruism</td>
<td>0:00:41</td>
</tr>
<tr>
<td></td>
<td>Civic Virtue</td>
<td>0:00:06</td>
</tr>
<tr>
<td></td>
<td>Courtesy</td>
<td>0:00:51</td>
</tr>
<tr>
<td>Peer Ratings</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>All</td>
<td>0:01:52</td>
</tr>
<tr>
<td></td>
<td>Altruism</td>
<td>0:01:27</td>
</tr>
<tr>
<td></td>
<td>Civic Virtue</td>
<td>0:01:37</td>
</tr>
<tr>
<td>Supervisor Ratings</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>All</td>
<td>0:02:59&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Altruism</td>
<td>0:27:43</td>
</tr>
<tr>
<td></td>
<td>Civic Virtue</td>
<td>0:11:38</td>
</tr>
<tr>
<td></td>
<td>Conscientiousness</td>
<td>0:05:44</td>
</tr>
<tr>
<td></td>
<td>Courtesy</td>
<td>0:03:25</td>
</tr>
<tr>
<td></td>
<td>Sportsmanship</td>
<td>0:03:15</td>
</tr>
</tbody>
</table>

Note. The times listed are expressed in hours: minutes: seconds. <sup>a</sup>This time reflects analysis of a model in which the maximum number of groups per organization, and maximum number of individuals per group was limited to 10. A model was also run where this maximum number was allowed to increase to 16, that model took 2:11:20 to converge to a solution. <sup>b</sup>This time reflects analysis of a model in which the maximum number of groups per organization, and maximum number of individuals per group was limited to 10. Attempts to run this model where this maximum model was allowed to increase to 16 failed (not enough memory). Runs were made using SAS-STAT 8.0’s Proc Mixed procedure. The software was installed a 20 GB partition, of a Pentium III 900 Mhz PC, with 128 MB of RAM.

approach is a large-scale technique. The present data represent an extremely small portion of what would be necessary to fully delineate the architectures underlying the OCB construct, both in terms of the sheer amount of data needed as well as the complexity of the designs examined. Given the estimation problems experienced with the small amount of data gathered for this study for relatively simple designs, such problems are likely to exponentially increase when more data and more complex architectural designs are examined.

As discussed in Appendix A, one of the potential ways to work around this limitation in future efforts involves repeatedly drawing smaller samples from the data available and essentially generating bootstrapped estimates of all parameters of interest (e.g., \( \sigma^2 \) and \( p_{\sigma^2} \) for each facet as
well as their confidence intervals; Efron & Tibshirani, 1986). However, even if such a repeated sampling method was used, it would still likely require supercomputing power due to the fact that the “smaller” samples would still likely be much larger than the sample investigated in the present study and the designs examined much more complex.

Another alternative would be to investigate other forms of variance component estimation besides that of REML that retain the optimal properties that make REML estimates currently in favor in the statistical literature (Searle et al., 1992). Unfortunately, such alternative methods for estimating variance component are still in their early stages of development (e.g., RAVE or AVE-based estimators; Hocking, 1990; Mehrotra, 1997, see Appendix A).33 Nevertheless, by the time enough data is collected for another round of VAA-related analyses, enough developments in the statistical literature may occur with regard to these alternative estimators to make them worthy replacements for REML.

Refining Methods for Comparing Variance Architectures

Another challenge for the future application of the variance architecture approach is to develop methods for inferentially comparing indexes of the architecture-level similarity such as the ASI-RC and ASI-RS values introduced in the present study. One of the weaknesses of these indexes is that they do not have known sampling distribution, thus, it was not possible to make statements regarding whether (a) observed differences between these values were statistically significant or (b) high values indicated a statistically significant degree of similarity among architectures. Indeed, as described earlier, these methods simply served as descriptive indexes that provided indicators of the overall similarity among architectures. Further work with regard to these indexes is required so that in the future one can provide more rigorous tests of architecture-related hypotheses.
Although, the current scarcity of inferential methods available for making many of the architecture-level comparisons raised above is unfortunate, it is likely that traditional tests of significance may become quite meaningless once the data upon which construct’s variance architectures are being delineated becomes so large that even inconsequential differences become statistically significant. Many in the field have criticized this potential to "over-power" inferential tests as a weakness of the traditional significance testing, and it may be particularly true for the VAA, given it is a large-sample methodology (Hunter & Schmidt, 1990; Rosenthal, 1983). The dilemma within the VAA may be that when samples become very large, the comparisons discussed above may be more a matter of directly comparing sets of population parameter, rather then inferentially comparing sets of sample statistics. When this point is reached, the pertinent question is not one of statistical significance, but how large of a difference would need to be observed between variance architectures or their parameters to suggest noteworthy differences exist (i.e., critical effect sizes).

One potential answer to the question posed above is to generate normative values for contributions expected for each facet across all constructs in general and their relative stability. For example, one could calculate the standard deviation of the contribution of the individual facet across all constructs for which variance architectures have been delineated and judge the relative magnitude of the individual facet for any particular construct expressed in terms of the number of standard deviations it deviates from the mean contribution of the individual facet across all constructs. Such a method could be used to identify constructs that are quite similar with regard to the contributions made by a given facet (e.g., within a tenth of a standard deviation of each other) or quite different (e.g., more than a standard deviation apart). Of course, the criteria for determining what constitutes, small, medium, or large differences (i.e., a critical number of standard deviations) would have to be established based on more substance than this example
provides, nevertheless this example illustrates how one might compare architectures once population-like values for many constructs' variance architectures are achieved.

**Next Steps**

Although many of the issues above limit the immediate use of the variance architecture approach as it is formulated in this paper, there are some aspects of the VAA methodology that lend themselves to have more immediate impact. As the example below will illustrate, the variance decomposition aspect of the VAA in particular can offer immediate benefit and insight to organizational practitioners.

Consider an organizational practitioner who wishes to design and implement a program to enhance the job satisfaction of the workforce in the organization for which he or she works. As a first step in this process the practitioner could use the variance decomposition procedures underlying the VAA approach to delineate a variance architecture underlying job satisfaction data gathered in an organization-wide survey that is extremely detailed with many facets specific to his or her particular organization. What the variance decomposition would do at this stage of development in an intervention effort would be to pinpoint for the practitioner the most pertinent sources of variation in individuals’ job satisfaction throughout the organizations based on a function of the facets included in his or her architecture, which would be based on the hierarchy specific to that given organization (e.g., regions, divisions, departments, workgroups, job families, jobs etc.). Based on the knowledge gained by delineating such a detailed architecture for job satisfaction in his or her organization, the practitioner would have a solid idea regarding where he or she can potentially affect the largest sources of variation in the job satisfaction.

Upon identifying theory-based interventions designed to effect variance at the levels at which it is most occurring and implementing the intervention, the practitioner may then conduct a follow-up job satisfaction survey and delineate a second architecture for job satisfaction based on
the post-intervention data. Should the intervention account for a substantial amount of variance at the level at which it was designed to function, then the post-intervention architecture should reveal a decrease in the amount of variance due to that facet and an overall reduction in the total variance of satisfaction across the organization, as well as an increase in the grand mean of satisfaction across the entire organization. For example, if the pre-intervention delineation of the architecture underlying job satisfaction revealed that department membership accounted for a large portion of variance in individuals’ job satisfaction, the practitioner might focus on identifying what department-level variables or conditions are at the root of such variation and design interventions that increase satisfaction in the less satisfied units, thus reducing variation through attempting to raise the “floor” for the satisfaction variable in departments suffering from low satisfaction. Following this intervention effort targeted toward specific departments, the practitioner could then conduct a follow-up survey and decompose the post-intervention data to check for an effective decrease in department-level variance and an increase in the overall satisfaction of the organization. This process can be repeated to continually identify and target the largest sources of variation in individuals’ levels of job satisfaction and ensure the variation due to other facets in the architecture remains stable.

As the above example illustrates, the variance decomposition aspect of the VAA can be immediately useful to OD practitioners due to its (a) allowing for the flexible designation of sources in the design based on the particular organization's hierarchical structuring, (b) pinpointing sources of variation in phenomena of interest to OD researchers for uses in a diagnostic phase of intervention efforts (thus saving time and money by not targeting levels that account for little of the variation in the organizational phenomena as a whole), and (c) acting as a tool for which the effects of the change efforts can be evaluated and rechecked for stability. Teamed with the research-oriented perspective offered in this paper’s exposition of the VAA,
applied work such as this can help establish a strong case for broad acceptance and use of the methodology described in this paper.

A Final Note

The key to realizing the benefits offered by the variance architecture approach is cooperation. Cooperation first and foremost among academicians in the variety of disciplines that each contribute to the body of organizational research (e.g., education, psychology, management etc). Cooperation, second in terms of the collaboration between academicians and practitioner's of organizational science in the field, as VAA-related data is extremely difficult to generate without well-maintained ties between academics and their colleagues in the field (Markowitz, Guzzo, Mastrangelo, Offerman, & Vasilopoulos, 2000). Although some of the methods discussed in this study can provide benefits if applied within single organizational contexts (e.g., for OD purposes), to fully realize the utility of the VAA would require a joint effort among organizational researchers.

Achieving a fully detailed description of a construct’s variance architecture would likely necessitate having a centralized database to which data may be contributed on an ongoing basis and a construct’s architecture subsequently refined based on newly acquired data. Successfully establishing and growing such databases can only be achieved through a unified effort. Unfortunately, there are many mechanisms at work in both academia and practice that reinforce individual efforts on the part of researchers rather than extending joint efforts for the common good of the field (Azar, 1999). As such, until new mechanisms are created to reinforce or encourage the cooperation, the vision laid out in this paper will be exceedingly difficult to achieve.
References

References marked with an asterisk indicate studies included in the initial sampling frame. Those with a double asterisk indicate those that provided data sets for inclusion in the present study.


Appendix A: Software and Hardware Issues Considerations in the VAA

As alluded to in the introduction, generating REML-based estimates of GLMM parameters is an extremely computationally intensive procedure (Littell et al., 1996; Searle et al., 1992; SAS Institute Inc., 1999). Given the large amount of data the VAA will focus on, software and hardware issues are a central concern. For small amounts of data, generating REML estimates of variance components can be achieved via commonly available commercial software programs (e.g., SAS). As suggested in the methods section, one of the more flexible, and easy to use choices is available through SAS's Mixed procedure (Littell et al., 1996; SAS Institute Inc., 1999). SAS's "Proc Mixed" provides a flexible and powerful means for estimating variance components using a variety of potential estimation methods (e.g., EMS, MIVQUE0, ML, REML), for any variety of random and mixed designs. Moreover, Proc Mixed allows one to fit a variety of covariance structures for D and R, and allows one to compare the relative fit of models based on different covariance structurings. For example, when one is examining occasion as a facet, and wishes to compare the relative fit of models that specify traditional variance component structures versus those that specify autoregressive-type structures for R, one can easily do so within the Mixed procedure (e.g., Littell et al., 1996).

Although SAS's Mixed procedure is easy to work with and quite flexible, it may not be the ideal solution for generating variance component estimates within the VAA due to the large amount of data it must process. This is not necessarily a limitation of the SAS software, but rather a combination of the hardware a typical researcher would run it off of (e.g., a standard desktop computer), as well as the computational requirements of the algorithms underlying the generation of REML-based variance component estimates themselves. To elaborate further, SAS's technical documentation provides formulas for approximating the minimum amount of memory required for generating REML variance components based on the amount of data one is examining. Specifically, one can compute the minimum memory required to generate REML-based variance components using Proc Mixed with the following formula:

\[
(40(1+q^2) + 32(1+q)^2)/1000
\]

In this equation, \( q \) refers to the total number of non-residual random effects in the model (i.e., the number of columns in the Z matrix). The result of this formula is the minimum amount of memory in kilobytes that is required by SAS to generate REML-based estimates of the variance components for a given model.

Recall from Example 4 in the text (examples of the variance decomposition process for different designs), the data collection design was such that information on two behaviors were gathered for eight individuals who were nested within four workgroups that were nested within two organizations. The total number of columns of Z for this data set was 28 (2 item random effects, 4 random effects stemming from the organization-by-behavior interaction, 8 stemming from the group-by-behavior interaction, 2 organization random effects, 4 group random effects, and 8 individual random effects). Based on Equation A1, the minimum amount of memory required to generate the variance component estimates using SAS would be only 58.32 kilobytes. Within the variance architecture approach however, one would be using far more data, and eventually much more complex designs.
To illustrate the problem large amounts of additional data raises, let's say the design from Example 4 remained the same, but the sample sizes for each facet increased substantially. For example, let's say that one now had data for on 10 behaviors for 10,000 individuals nested within 1,000 workgroups, which were drawn from 100 organizations. The total number of random effects for this amount of data would now jump to 22,110 (10 item random effects, 1,000 random effects stemming from the organization-by-behavior interaction, 10,000 stemming from the group-by-behavior interaction, 100 organization random effects, 1,000 group random effects, and 10,000 individuals random effects). The minimum amount of memory now required to generate the variance component estimates using SAS would jump to 351,987,663 kilobytes (i.e., 351 gigabytes). The primary memory burden comes from having to store symmetric matrices of the order $q$ and $q + 1$, where $q$ in this example is 22,110.

The complication such memory requirements create are substantial for analyses run on a standard desktop computer, because desktop hard drives of such computers rarely exceed 100GB, and common desktop operating systems (e.g., Windows) often put a hard limit on the amount of virtual memory that can be allocated to any given program (e.g., 4GB). Given the latter point, even if the hard drive storage capacities were substantially increased, limitations would likely be imposed by the desktop's OS to maintain the stability of the rest of the system. Although desktops suffer from such drawbacks, running SAS off of larger computers, namely university servers or regional supercomputers may provide for a much more satisfactory solution. The amount of virtual memory that can be allocated to batch statistical jobs of the nature described above are far less limited on such computers. Nevertheless, even if one were able to acquire time on a regional supercomputer, the dollar cost of doing so may be substantial given the estimated computing time required by the algorithms that REML-based estimates require (see SAS Institute Inc., 1999, chapter 41- computational issues, for more details).

Ultimately, the solution to the aforementioned computational problems with regard to providing statistically optimal estimators of variance components within the VAA may be resolved via several possible routes. One potential solution would be to work within a supercomputing environment, and repeatedly draw smaller samples from a larger sample of VAA related data in order to generate multiple REML-based estimates for the variance components. Not only would this greatly reduce the computational burden required to estimate any single "run" of the model, but it would also allow one to generate bootstrapped standard errors for variance components that would arguably be highly robust to any departures of normality that limit traditional variance component standard error estimates (e.g., Arvesen & Layard, 1975; Brennan, 1994). Arguably, not a great deal of information would be lost by creating such smaller random samples because they would still be much larger than what is typically viewed as required for generating stable estimates for variance components (about 10-15 units per facet; Smith, 1978). A potential problem with this strategy would be determining how to randomly sample data from complicated multi-facet designs, as one would be sampling across more than just a single facet (e.g., individuals) which is often the focus of bootstrap-based methods (e.g., Efron & Tibshirani, 1986).

Another potential solution to the computational problems surrounding the generation of REML-based variance components for large data sets would be to explore other emerging methods of variance component estimation that retain their optimal statistical properties across a variety of conditions (e.g., unbalanced designs), yet are far less computationally intensive. Although work is being done within the statistical literature to these ends, most of the efforts are
in their preliminary stages (e.g., Green, 1988; Hocking, Green, & Bremer, 1989; Hocking, 1990; Mehrotra, 1995, Mehrotra, 1997). One particular estimator that appears promising is known as the RAVE estimator (robust AVE estimators; Mehrotra, 1997). RAVE estimators extend Hocking's AVE method of estimating variance components that is based on weighted averages of sample variances and covariances to obtain variance component estimates (Hocking et al., 1989; Searle et al., 1992). Unlike, Hocking's AVE estimators, however, RAVE estimators are generally robust to departures from normality, and appear appropriate for both balanced and unbalanced designs (Mehrotra, 1997). In the introduction of RAVE estimators to the literature, Mehrotra (1997) conducted simulations that illustrated how these estimators tended to retain their optimal statistical properties with unbalanced designs, are more robust to violations of normality (with regard to random effects) and are far less computationally intensive (no iteration, or numerical analysis required) than REML estimators. Unfortunately, there has been little published follow-up work based on Mehrotra's (1997) RAVE estimators, thus, implementing such estimators prior to their further refinement in the statistical literature may be suspect.
Appendix B: Formulas for Confidence Intervals around Facet Contributions

Relative Facet Contributions

Upon generating estimates for the proportions of variance accounted for by facets of the variance architectures examined, (i.e., $p_{\sigma^2}$), confidence intervals were formed around them using the following series of formulas stemming from Fisher’s (1925) ICC to $z$ transformation. First, the $p_{\sigma^2}$ corresponding to each facet in a design was transformed to a normal $z$ deviate using Fisher's (1925) ICC to $z$ transformation, which is as follows:

$$ z = .5 \log \frac{1 + (k - 1)r}{1 - r} $$

(B1)

Recall, that the proportions of total variance accounted for by a facet can be viewed as specific types of ICCs (McGraw & Wong, 1996; Shrout & Fleiss, 1979). Thus, to transform the $p_{\sigma^2}$ values to normal $z$ deviates, $p_{\sigma^2}$ can simply be substituted for $r$ in Formula B1. The $k$ in this formula becomes the number of observations available for each unit of the facet of interest.

One of the benefits of transforming the $p_{\sigma^2}$ values to $z$ using Fisher’s formula is that it also results in an estimate for the standard error of $z$, which is:

$$ S.E.(z) = \frac{k}{\sqrt{2(k - 1)(n - 2)}} $$

(B2)

Where $k$ is as stated above, and $n$ is the number of units sampled from the facet of interest. In cases where an unequal number of observations was available for each unit (e.g., individuals differing in the number of responses they provided; the median number of observations available across each unit was used as an estimate for $k$.

Both $z$ and $S.E.(z)$ were calculated for each facet. Upon generating these values, a 90% confidence interval was formed around the $z$ value for a given facet by using the following formula:

$$ 90\% \text{ C.I.}: z - 1.65(S.E.(z)), z + 1.65(S.E.(z)) $$

(B3)

Once the upper and lower confidence interval bounds on the normal $z$ deviate were established for each facet, these bounds were transformed back to the raw $p_{\sigma^2}$ metric using Fisher's (1925) $z$ to ICC transformation. The formula for converting these values back is:

$$ r = \frac{1}{2(k - 1)} \left( k \left( \frac{e^{2(z - .5 \log(k - 1))}}{e^{2(z - .5 \log(k - 1))} + 1} \right) - 1 \right) + (k - 2) $$

(B4)

All components of this equation are as indicated above. Transforming these values using this formula results in a 90% confidence interval around $p_{\sigma^2}$ in the raw $p_{\sigma^2}$ metric. Given that $p_{\sigma^2}$
‘s are not normally distributed, the confidence intervals surrounding the $p_{\sigma^2}$ ‘s (unlike those surrounding their corresponding normal z-deviates) are asymmetric.

**Absolute Facet Contributions**

Upon generating estimates for the variance components corresponding to each facet one can form approximate confidence intervals around those estimates using Satterthwaite’s procedure (1946). Based on Satterthwaite’s procedure, the 90% confidence interval surrounding a given variance component can be approximated by the following formula:

$$90\% \text{ C.I.}: \left( \frac{v_x}{\chi^2_{.95}(v_x)} \right) \cdot \sigma_{x^2} \left( \frac{v_x}{\chi^2_{.05}(v_x)} \right)$$  \hspace{1cm} (B5)

where $v_x = 2 \left( \frac{\sigma_x^2}{\text{S.E.} \sigma_x^2} \right)^2$  \hspace{1cm} (B6)

In the above equation, $\sigma_x^2$ is the variance component of the facet of interest ($x$). $v_x$ is the “effective degrees of freedom” for estimating the confidence interval around $\sigma_x^2$, and is calculated via Formula B6. The $\text{S.E.} \sigma_x^2$ in B6 is the asymptotic standard error estimate for the given variance component generated by most statistical programs that estimate variance components (e.g., SAS, SPSS). The $\chi^2$ value in B5 is the value of chi-squared distribution corresponding to the given cumulative density (i.e., .95 or .05) with $v_x$ degrees of freedom. As cited earlier, the confidence interval for variance components produced by this formula will tend to be asymmetric because estimates of variance generally have a positively skewed sampling distribution (although such a distribution becomes more normal as the number of units sampled for a given facet grows large).
Appendix C: Formulas for Estimating the Stability of Facet Contributions

To estimate the stability of $p_{\sigma}^2$ values for a given facet across organizations, an i:g x b design was estimated separately for each organization (for each OCB-dimension-by-rating method combination separately). Estimating the stability of the $p_{\sigma}^2$ values that resulted for the facets in each of these architectures across organizations was achieved using meta-analytic methods derived from Hunter and Schmidt’s (1990) techniques. To estimate the stability of $p_{\sigma}^2$ for a single facet across organizations, the series of steps outlined below were undertaken.

First, the $p_{\sigma}^2$ values for the facet generated in the $k$ analyses of the i:g x b (one for each organization) were transformed to their corresponding normal z-deviates using Fisher’s (1925) methodology described in Appendix B. At the end of this step, there were $k$ normal z deviates, and $k$ standard errors associated with these normal z deviates that served as the data to be meta-analyzed. Most of the meta-analytic operations are carried out on the z normal deviates as opposed to the raw $p_{\sigma}^2$ values because the z normal deviates have an associated standard error, which is critical for meta-analytic work.

The next steps in the process were to (a) meta-analytically aggregate the normal z deviates for a given facet, (b) generate estimates of their raw variation across organizations, and (c) estimate the amount of this variation that could be attributed to sampling error. The formula used to estimate the meta-analytic mean $z_{\sigma}^2$ (mean normal z deviate across organizations) was as follows:

$$Mz_{\sigma}^2 = \frac{\sum_k \left( \frac{1}{S.E.(z_{\sigma}^2_k)}z_{\sigma}^2_k \right)}{\sum_k \left( \frac{1}{S.E.(z_{\sigma}^2_k)} \right)}$$

(C1)

The primary difference between this formula for the meta-analytic mean and the formula offered by Hunter and Schmidt (1990) is that the sample statistics (normal z deviates in this case) are weighted by the inverse of their standard errors instead of the size of the samples from which they were drawn. This is due to the fact that unlike traditional zero-order Pearson correlations, the standard error of an ICC is more than just a function of sample size (namely, it is also a function of the number of observations sampled for each unit of the facet of interest- refer to Formula B2 in Appendix B). Thus, the contribution of each normal z deviate to the mean depended on the precision with which it was estimated in the organization in which the value was generated.

Next, the raw variance across these normal z deviates across organizations was calculated using the following formula:
Now given that part of this “raw” variance likely reflects artifactual variation arising from sampling error within each organization, an estimate of the variance in normal z deviates due to sampling error was generated using the following formula:

\[
Var(z_{\sigma}^2)_{\text{Raw}} = \frac{\sum_k \left( \frac{1}{S.E.(z_{\sigma}^2)} (z_{\sigma}^2 - Mz_{\sigma}^2)^2 \right)}{\sum_k \left( \frac{1}{S.E.(z_{\sigma}^2)} \right)}
\]

Subtracting the result of formula C3 from the result of formula C2, gives an estimate of the “true” variance in a facet’s relative contribution to construct variance across organizations after removing the variance due to sampling error. Based on Hunter and Schmidt's (1990) 75% rule, if more than 3/4 of the raw variance (calculated in C2) can be attributed to sampling error, or other artifacts, than it is likely that no meaningful moderators of the magnitude of the given facet’s effects exists across organizations, and that the remaining variation can likely be attributable to other artifacts that were not considered. In the present study, the ratio of error variance over raw variance was compared to see if it met this standard. Along with this 75% rule of thumb, the \(Q\)-test for homogeneity (outlined in the text) was also calculated to provided the test of whether or not the contribution of a facet significantly varied across organizational settings, or whether such variation likely just arose from as an artifact of sampling error.
Appendix D: Initial Letter Sent to Researchers for Raw OCB Data

Dear Dr. __________,

My name is Dan Putka and I'm a doctoral candidate in the Industrial-Organizational psychology program at Ohio University. I am writing you today because I'm interested in reanalyzing a small portion of Organizational Citizenship Behavior (OCB) data from your year, journal article titled "___________" as part of my dissertation. The following is a very brief overview of what my dissertation is about, why your data would be helpful, and what data I am requesting from you.

An Overview of My Dissertation

For my dissertation, I will be estimating the potential contribution of factors stemming from several "levels" of organizational functioning to the total variance in employees' OCB engagement. My main premise is that OCB is often viewed as an individual level construct, and that makes it easy to lose sight of the fact that variance in OCB stems from variables that arise at several different levels (e.g., the individual, job, group, departmental, and organizational levels). I argue that estimating the degree to OCBs vary across such levels of the organizational hierarchy can be very helpful in several regards for guiding both organizational science and practice.

Why Your Data Would Be Helpful

Accomplishing the goals outlined above will require the acquisition of several "raw" OCB data sets from many different OCB researchers (I will need OCB data from a wide variety of individuals, groups, organizations etc., in order to capture variance across the levels I'm examining). Given the likelihood that one study will not contain data from a wide variety of organizations, I'm currently in the process of contacting basically every other published OCB researcher for their raw data sets as well. The data from your study will be a great contribution to my dissertation and will help ensure that the sample I eventually obtain is representative of the OCB literature.

What Data I am Requesting from You

It is important to realize, that in asking you for your data the only variable I'm interested in are the ratings of employees' engagement in each of the OCBs that you published in the article cited earlier (or any other OCB data you might have collected that fit the characteristics described below). I don't need any data on any other variables you collected (i.e., no specific predictor variables). The only other request I'm making regards some basic information on how, where, and from whom this data was collected. For example, I would like to identify the general type of organization the data was collected from (e.g., small Midwestern manufacturer), the type of ratings that were used (e.g., self, supervisor, peer), what raters rated which employees, the specific OCBs you examined, and the jobs the target employees held (For an exact summary of what I am requesting, as well as other specifics, see the postscript at the end of this message).

I realize that this information may be hard to obtain, and the files that contain this information hard to find. Although what I'm proposing is quite an undertaking, the gains of doing this are substantial, and I feel that your contribution to this project is very important.

My name is Dan Putka and I'm a doctoral candidate in the Industrial-Organizational psychology program at Ohio University. I am writing you today because I'm interested in reanalyzing a small portion of Organizational Citizenship Behavior (OCB) data from your year, journal article titled "___________" as part of my dissertation. The following is a very brief overview of what my dissertation is about, why your data would be helpful, and what data I am requesting from you.

An Overview of My Dissertation

For my dissertation, I will be estimating the potential contribution of factors stemming from several "levels" of organizational functioning to the total variance in employees' OCB engagement. My main premise is that OCB is often viewed as an individual level construct, and that makes it easy to lose sight of the fact that variance in OCB stems from variables that arise at several different levels (e.g., the individual, job, group, departmental, and organizational levels). I argue that estimating the degree to OCBs vary across such levels of the organizational hierarchy can be very helpful in several regards for guiding both organizational science and practice.

Why Your Data Would Be Helpful

Accomplishing the goals outlined above will require the acquisition of several "raw" OCB data sets from many different OCB researchers (I will need OCB data from a wide variety of individuals, groups, organizations etc., in order to capture variance across the levels I'm examining). Given the likelihood that one study will not contain data from a wide variety of organizations, I'm currently in the process of contacting basically every other published OCB researcher for their raw data sets as well. The data from your study will be a great contribution to my dissertation and will help ensure that the sample I eventually obtain is representative of the OCB literature.

What Data I am Requesting from You

It is important to realize, that in asking you for your data the only variable I'm interested in are the ratings of employees' engagement in each of the OCBs that you published in the article cited earlier (or any other OCB data you might have collected that fit the characteristics described below). I don't need any data on any other variables you collected (i.e., no specific predictor variables). The only other request I'm making regards some basic information on how, where, and from whom this data was collected. For example, I would like to identify the general type of organization the data was collected from (e.g., small Midwestern manufacturer), the type of ratings that were used (e.g., self, supervisor, peer), what raters rated which employees, the specific OCBs you examined, and the jobs the target employees held (For an exact summary of what I am requesting, as well as other specifics, see the postscript at the end of this message).

I realize that this information may be hard to obtain, and the files that contain this information hard to find. Although what I'm proposing is quite an undertaking, the gains of doing this are substantial, and I feel that your contribution to this project is very important.
Thank you for taking the time to consider my request and I look forward to hearing from you soon. The best way to contact me is via e-mail, however, if you prefer to talk to me by phone, you can call me at (412)-257-5379 during the day, or (412)-220-9644 in the evening. If I don't hear from you in a couple weeks, I'll try reaching you by phone to confirm your receipt of this message.

Sincerely,

Dan J. Putka
Industrial-Organizational Psychology Program
200 Porter Hall
Ohio University
Athens, OH 45701
Daytime Phone: (412)-257-5379
Evening Phone: (412)-220-9644
e-mail: dan.putka.1@ohiou.edu

P.S. A Detailed Description of the Data I am Requesting

A description of the data I am requesting, as well as issues surrounding its formatting, and a timeline for my request can be found below. For the data I am requesting, I have organized what I am requesting by how essential the information I am seeking is to my project.

Essential Information

1. Ratings of OCB engagement for each employee at the level of the individual behavior (i.e., not aggregated across behaviors into dimensions or an overall OCB rating). This is very important-- if the files don't contain ratings of individual behaviors, I won't be able to use them.

2. The type of organization each data set came from (e.g., small Midwestern steel manufacturing plant, hospital etc.).

3. Which supervisors rated which employees, or if peer or self-ratings were used, which employees worked under which supervisors (e.g., supervisor A rated employees 1 and 2, supervisor B rated employees 3 and 4, or Employees 1 and 2 worked under Supervisor A etc.).

4. The specific OCBs examined (e.g., OCB1 is "helps co-workers who have heavy workloads", OCB2 is "attends meetings" etc.).
5. The "provider" of the OCB ratings (e.g., supervisor, peer, or self-ratings).

6. The type of jobs the target employees held (e.g., employee 1 was a registered nurse, employee 2 was a clerical worker, employee 3 was a manager). I realize the level of detail here may vary-- whatever level of detail you have recorded will be fine.

Helpful, yet not essential information

1. Any departmental or regional distinctions that could be made between employees would be great. This becomes more relevant in large organizational samples.

2. The source of data (published article, presentation etc.). Although I'm primarily requesting data from your published article(s), this distinction may be in order if you have other OCB data you might be willing to share.

3. If any OCB data you have is longitudinal, I would need to know how much time occurred between the ratings you collected, and be able to discern the order in which the ratings took place.

Formatting Issues

How you have the data saved is not important (e.g., in multiple files-- one for each organization / study, saved in Excel, SPSS, SAS, or ASCII format). I will be responsible for getting the files into the format I need, and merging files together where possible. What is important is that I can identify the various aspects of the OCB ratings mentioned above. If you have the above information, but it is not in the data files, that is also fine. For example, some of this information may be available in a "codebook" or syntax file you created to accompany your data files. No matter what the format, any documentation you have that would help me identify the information I requested above would be helpful.

Time Constraints

As far as a time line on the above request, I'm very flexible. I realize you're likely very busy, and finding the aforementioned file(s) may take some time. If possible I would like to try to obtain any files you have to offer by date (four weeks after this initial request was sent). If that is too soon let me know.

If you agree to aide me in my endeavor by supplying me with OCB data you have collected, you can send me the data files and codebooks/syntax files (if you have them) as attachments over e-mail. If you feel uncomfortable with this, we can make other arrangements. Thanks again for your help!
Appendix E: Follow-Up Letter Sent to Researchers for Raw OCB Data

Dear Dr. __________,

Back on date initial request was sent, I e-mailed you a request for OCB data that I'm seeking for my dissertation. As I have yet to hear from you, I was just wondering if you had time to consider my request. For your convenience, I have pasted my original request in the postscript of this message below.

If you do not feel you can provide me with this data I have requested, please respond to this message and let me know. I am trying to keep track of my contact with over 100 researchers, and if you cannot help me at this time, your response to this mailing will prevent me from sending you follow-up requests in the future. Thanks again for your time!

Sincerely,

Dan J. Putka
Industrial-Organizational Psychology Program
200 Porter Hall
Ohio University
Athens, OH 45701
Daytime Phone: (412)-257-5379
Evening Phone: (412)-220-9644
e-mail: dan.putka.1@ohiou.edu

P.S. This postscript contained the letter that appears in Appendix D.
Appendix F: List of OCB Scale Items Examined by Dimension

**Altruism**

Helps others who have been absent.
Helps others who have heavy workloads.
Helps orient new people even though it is not required.
Willingly helps others who have work-related problems.
Is always ready to lend a helping hand to those around him/her.

**Civic Virtue**

Attends meetings that are not mandatory, but are considered important.
Attends meetings that are not required, but help the company image.
Keeps abreast of changes in the organization.
Reads and keeps up with organization announcements, memos, and so on.
Provides constructive suggestions regarding changes that might be made in his/her department or the company.
Is willing to risk disapproval in order to express his/her beliefs about what's best for the department/company.

**Conscientiousness**

Has attendance at work that is above the norm.
Does not take extra breaks.
Obeys company rules and regulations even when no one is watching.
Is one of my most conscientious employees.
Believes in giving an honest day's work for an honest day's pay.

**Courtesv**

Takes steps to try to prevent problems with other workers.
Is mindful of how his/her behavior affects other people's jobs.
Does not abuse the rights of others.
Tries to avoid creating problems for coworkers.
Considers the impact of his/her actions on coworkers.

**Sportsmanship**

Consumes a lot of time complaining about trivial matters.
Tends to make "mountains out of molehills".
Always finds fault with what the organization is doing.
Is the classic "squeaky wheel" that always needs greasing.
Appendix G: Formulas for Estimating Random Effects of Each Facet

Estimating the potential for non-normality in the distributions of random effects stemming from each variance architecture facet was achieved by generating sample estimates of the random effects and then checking their resulting distributions for departures from normality using indexes of skewness and kurtosis. Sample estimates for the random effects were generated separately for each OCB-dimension-by-rating method combination using the formulas outlined by Brennan (1983) in his exposition of Generalizability Theory (see Brennan, 1983, pps. 9 and 92).

Generating estimates of random effects first requires that one specify a particular design by which variance in the dependent variable of interest will be decomposed. The primary design of interest in this study was i:g:o x b, and thus the focus of the normality checks centered around the facets stemming from that design. The dependent variable in this design is individuals’ engagement in OCB. One can express the observed values of individuals’ engagement in specific OCBs as a function of the random effects stemming from the i:g:o x b design. Specifically,

\[
Y_{ijkl} = \mu + (\mu_b - \mu) + (\mu_g \circ - \mu_g) + (\mu_i \circ \circ - \mu_i \circ) + (Y_{ijkl} - \mu_i \circ \circ - \mu_g \circ) \quad \text{(G1)}
\]

In this equation, \(Y_{ijkl}\) is the observed engagement of individual \(i\) nested in group \(j\) nested in organization \(k\) in the \(l^{th}\) OCB. The \(\mu_{\text{facet}}\) values listed are unique for each unit of a given facet and thus these values will differ depending on the particular unit of the facet being considered. These values indicate the expected value of a unit of the facet in question on OCB engagement, across units of all other facets in the design. For example to generate a \(\mu_{i \circ \circ o}\) value for a unit of the individual facet (i.e., a single individual) one would average that individual’s engagement in OCB across all items on which he or she was rated. This average would serve as an estimate of the individual’s \(\mu_{i \circ \circ o}\) value across all possible OCBs he or she could be rated on (i.e., the population of items, of which we only have a sample- thus it’s only an estimate). Similarly, to generate the \(\mu_{g \circ o}\) value for a given unit of the group facet (i.e., a specific group), one would average across all of its members’ engagement in OCB on all OCB items on which they were rated. This average
would serve as an estimate of the group’s $\mu_{g,o}$ value across all group members and OCBs on which they could be rated (i.e., the population of individuals within that group and the population of items, from both of which we only have a sample of their respective units).

Upon generating these mean values for each unit of each facet in the architecture, the random effects associated with each unit of each facet were generated by subtracting and/or adding the appropriate mean values as indicated in equation G1 above. Carrying out these operations resulted in sample distributions of individual effects, group effects, organization effects, behavior effects, organization-by-behavior effects, group-by-behavior effects, and individual-by-behavior effects (residual) for each dimension and each rating source. These distributions were then examined for normality by calculating skewness and kurtosis values for each of them.
Endnotes

1 Following the lead of Cronbach, Rajaratnam, and Gleser's (1963) in their introduction of Generalizability theory, I speak of “facets” of variation rather than “factors,” to avoid confusion with the literature on factor analysis.

2 To be more exact, it is not the impact that the common facets themselves have on the total variance in a construct, rather it is the impact that any single variable or set of variables stemming from those facets may have on the total variance in a construct.

3 Generalizability coefficients can be viewed as G-theory analogs of classical test theory’s reliability coefficients.

4 Actually, the statistical model underlying G-theory (known as the variance components model in the statistical literature; Searle et al., 1992) is a specific subclass of the more general class of models known as general linear mixed models. GLMMs in turn are a specific subclass of the more general class of models known as generalized linear models (McCullagh & Nelder, 1989).

5 While I acknowledge that random disturbance is still part of the measurement process of any subjectively assessed construct, within the VAA it is inseparable from the highest order interaction term in one’s model of variance in a construct. Taking this as a given, this random disturbance is not formally included in this exposition of the primary facets, as the highest order interaction term in any VAA model will be interpreted as a function of both a true interaction effect and residual variance.

6 Interestingly, Cattell (1966) does not recognize occasion as a facet within the BDRM. Nevertheless, he does recognize occasion-related variation. For example, within the BDRM, occasion-related variation stems from the interaction of occasion and each of the first five facets of the BDRM (e.g., state of the person, phase of the environmental background). In the VAA, these latter five BDRM facets are explicitly recognized as two-way interactions between occasion-facet and other facets in the VAA framework. For example, the BDRM’s “state of the person” facet is recognized in the VAA as the interaction between the individual- and occasion-facets. Another example is the BDRM’s “phase of the environmental background” facet. The VAA recognizes this facet as the interactions between each of the environmental facets and the occasion-facet. (i.e., the culture-by-occasion, organization-by-occasion, group-by-occasion, and job-by-occasion interactions).

7 Given that strict random sampling is not always feasible, the sampling practice underlying both the VAA and G-theory is based on the Bayesian principle of exchangeability of observations (de Finetti; 1964; Novick, 1976). Essentially, the principle of exchangeability refers to the notion that: "Even though conditions of a facet have not been sampled randomly, the facet may be considered to be random if the conditions not observed in the...study are equally acceptable to, or exchangeable with, the observed conditions" (Shavelson et al., 1989, p. 927).

8 Such a model is ideal, not because having a construct that requires more facets of variation to adequately describe its variance architecture is better, but rather, because having a model that includes all facets allows one to more confidently determine the number of facets needed to adequately describe that construct’s architecture. Indeed given the number of ways a construct can vary based on all of the facets underlying the VAA, as well as all of their n-way interactions, one hopes that some level of parsimony can be achieved in describing the variance architecture of a construct (i.e., many of the facets or their interactions contribute little or nothing to total construct variance).

9 This claim may depend on whether the observers collaborated over time when doing their ratings versus doing their ratings independently. If they did collaborate over time, the observers may calibrate their ratings
as they begin to understand each others decision polices more thoroughly and, in such a case, they may
develop an observer-pair-specific idiosyncratic rating tendency. If the observers made their ratings
independently, and did not have any interaction, it is unlikely that there would be a reliable observer-pair-
specific rating tendency that would emerge.

10 In matrix notation, capital letters typically denote matrices, and lower case letter represent column
vectors (i.e., matrices with only one column).

11 In reality one will have many more units of such facets than two. However, for purposes of illustration,
examples are provided with very few units per facet as possible to keep the matrices displayed simple.

12 A simple reason one does this is that the statistical programs that generate estimates for the parameters of
these models (e.g., SAS) will not decompose variance in a set of observations without a residual
component. Thus, R must be included in the model that is fitted.

13 Notation for identity matrices of n x n dimension (I) was introduced to save space. An identity matrix is
a matrix of n magnitude with 0’s in its off-diagonal elements and 1’s in its diagonal elements.

14 Because this matrix will be block diagonal, with an identical block for each organization, only the first
block was presented to save space, the block on the lower right diagonal will be identically structured and
have exactly the same elements as the block presented.

15 In the design discussed in this example (i:g: o x b), the individual facet is nested within the group and
organization facets, thus the individual-by-behavior interaction actually reflects the individual nested
within group nested within organization-by-behavior interaction (i.e., the highest order interaction in the
design). Recall that there are not multiple observations on each behavior for individuals in this design, thus
variance stemming from this interaction is completely inseparable from variance stemming from the
residual term (Brennan, 1983).

16 The matrix was too large to display.

17 Although no normality assumptions regarding the distribution of random effects are required to retain the
optimal qualities of EMS-based estimators, where violations of normality assumptions do become
problematic is when attempting to produce standard errors for EMS-based variance component estimates
(Searle et al., 1992).

18 The troublesome aspect of meta-analytically examining the stability of absolute variance component
estimates across contextual factor units stems from lacking a method for generating an estimate of the
amount of variance in variance component estimates across contextual factor units that can be attributed to
sampling error (Hoyt & Kerns, 1999). This prevents one from generating an estimate of the degree to which
absolute variance component values for a given facet “truly” vary across contextual factor units. Appendix
C illustrates how generating such estimates are possible if one examines relative facet contributions.

19 The “2/m” in this equation reflects the maximum possible mean absolute difference observable between
the relative contribution of the facets of two architectures based on a design with m facet and interactions.

20 For those studies where it was not clear whether employees’ organizational behavior was examined by
reviewing their abstracts alone, the full-length article was reviewed.

21 Although only one data set contained peer ratings of OCB, I was able to examine variance architectures
underlying these ratings because they were gathered on many individuals across a diverse array of
organizations.
In one data set, some individuals were rated by two peers. Nevertheless, two factors prevented the inclusion of an observer facet in architectures based on peer ratings. First, sampling only two levels of the observer facet for each individual ratee (and in many cases just one – i.e., one peer rater for each ratee) would not allow stable estimates of the contribution of the observer facet to be generated. Second, the awkward coding of peer raters in the data set made it difficult to reliably distinguish between specific peer raters.

Although I acknowledge that the observer facet is completely confounded with the group facet for supervisor ratings, and with the individual-facet for self and peer ratings (and thus the observer/group- and observer/individual-facet labels are somewhat arbitrary), I will refer to these sources of variation as being the group-facet (for supervisor ratings) and the individual-facet (for self and peer ratings). This decision was made in part to simplify the verbiage and notation used in subsequent discussion of study results, as well as the fact that research questions primarily reference individual- and group-facets rather than the observer facet. The limitations that the aforementioned confounding creates when attempting to answer the research posed as part of this study will be explicitly recognized in the discussion section.

Although lacking the ability to provide a unique estimate for the individual-by-behavior interaction may appear problematic, this issue will partially be addressed by fitting a model with dimension included as a factor to see if it reduces the effects of the residual sources of variance in this first general architecture that is fit to the data (i.e., i:g:o). If such a reduction does occur, it would indicate that part of the residual / individual-by-behavior variation could be accounted for by differences arising from individual’s differential engagement in the dimensions of OCB (i.e., an individual-by-dimension interaction).

Unlike the i:g:o architecture, I was unable to acquire enough computing power to estimate the parameters of the i:g:o x b architecture for each rating source across all behaviors (across all OCB dimensions). I only had the computing power to examine the i:g:o x b architecture for each rating source across behaviors within a given OCB dimension.

Standardizing based on the observed range would not be suitable in this study because if one standardizes across units within a facet it would eliminate variance stemming from higher-level facets.

For the present study I chose not to test for departures from normality using traditional inferential tests such as the Kolmogorov-Smirnov test (KS-test), which tests the null hypothesis that the population distribution from which the sample data was drawn is a normal distribution (Chakravarti, Laha, & Roy, 1967). Given the problems associated with such tests (e.g., miniscule departures from normality are deemed statistically significant when sample sizes are large, substantial departures from normality may be missed when sample sizes are small), I decided to use raw skewness and kurtosis values and rules of thumb adopted in the literature (e.g., Tabachnick & Fidell, 1996) to evaluate potential departures from the normality.

The only primary deviation from the typical use of Hunter and Schmidt's (1990) techniques was that instead of using sample sizes to weight each z value corresponding to the relative contribution of each facet (analogous to effect sizes in traditional meta-analysis), they were weighted by their standard error. Such a deviation is necessary to reflect that the fact that the precision of such values is more than just a function a single sample size (e.g., n). Specifically, the precision with which such values are estimated is a function of both the number of objects of measurement (n, analogous to the number of units sampled in the target facet of interest), and the number of observations on each object of measurement (k; Fisher, 1925; McGraw & Wong, 1996).

Sportsmanship ratings were reverse-coded so that higher values on items tapping this dimension of OCB correspond to high values on other dimensions of OCB (i.e., high levels of OCB engagement).
Although this has the potential to introduce sampling bias, to the extent that additional units are randomly sampled (assuming the first set of units were randomly sampled as well), sampling bias should be less problematic.

Percent reductions in the contribution of the individual-by-behavior / residual interaction to total variance was calculated by: (a) taking the percentage of variance accounted for by the residual in architectures without dimension included as a facet minus the percentage of variance accounted for by the residual in architectures with dimension included as a facet, (b) dividing the resulting quantity by the percentage of variance accounted for by the residual in facets without dimension included as a facet, and (c) multiplying the resulting quantity times 100.

Because peers only rated two dimensions, only one heterodimension-monomethod ASI-RC value was available for within-peer architecture comparisons. That ASI-RC value was .672 for the Altruism-Civic Virtue comparison.

As an ancillary part of this investigation, I also generated MIVQUE0-based variance component estimates for each analysis for which REML variance components are reported. Both the REML and MIVQUE0 estimation methods resulted in similar rank-orderings of the magnitude of facet contributions, however variance component values differed slightly for the two estimation methods with the MIVQUE0 method resulting in slightly higher contributions stemming from the residual facet. Such findings are consistent with the statistical literature (Swallow & Monohan, 1984).

Data from Tsui et al. (1997) was reanalyzed as part of this study. Its collection was originally funded by a National Science Foundation grant, number SES-8922123.

Generating estimates for all of these random effects was an extremely laborious process. For those who are interested, the author has created an SPSS syntax file that will greatly expedite the generation of these values. For a copy of the syntax, please contact the author.