

Complementary Meta-Analytic Methods for the Quantitative Review of Research: 2. An Extended Illustration

**Aurelio José Figueredo, Candace Jasmine Black,
and Anne Grete Scott**
University of Arizona

In Figueredo, Black, and Scott (this issue), we presented the rationale for a complementary meta-analytic method to accompany traditional effects meta-analytic procedures. Here, we provide an example using Contents Meta-Analysis so that readers can become familiar with the application of the method and the implications of its use. This illustration will be presented in two major sections. First, we will describe an empirical example of a meta-analysis on retention in higher education where a Contents Meta-Analysis was conducted. Then we will show how the information gained in the Contents Meta-Analysis may be applied to address issues of generalizability.

Keywords: Contents Meta-Analysis, Meta-Analysis, Higher Education, Research Paradigms, Methods

Empirical Example: Retention in Higher Education

To illustrate the utility of contents meta-analysis as a supplementary method to effects meta-analysis, we will review an example involving the factors that predict retention in higher education. Retention may be defined as the "ability of an institution to retain a student from admission through graduation" (Berger, Blanco Ramírez, & Lyons, 2012, p. 12). The analysis described herein constitutes an "ecological" assessment of attrition by looking at the personal and situational factors that influence the outcome. A second, less obvious, interpretation is that we are also assessing the "ecology" of the researcher, insofar as the patterns in measurement reflect characteristics of investigators and their approaches to conducting research.

Assessing the impact of these factors is important because completion of a university degree program is associated with a number of positive benefits for the student, including both improved academically-related skills as well as general life skills. With regard to academic skills, freshman-to-senior gains have been demonstrated in a number of areas, including verbal and quantitative skills, speaking and written communication, critical thinking, and reasoning. Socioeconomic outcomes are improved as well. College graduates, compared to high

school graduates, are able to obtain more desirable jobs and enjoy increased earnings. Evidence also shows that college graduates experience higher psychological well-being, independence, and improved interpersonal relations (Pascarella & Terenzini, 2005). Furthermore, the student is not the only beneficiary of a college education. Indeed, there are a number of stakeholders in the higher education enterprise, including the institution, educators, parents, taxpayers, federal, state, and local governments, and, arguably, society at large. Given this collective investment in the next generation, assessing the factors that facilitate (or hinder) completion of a college degree seem warranted.

Methods

Empirical studies published between 1949 and 1992 were retrieved for this study. The initial search produced 240 references which were evaluated for relevance to this meta-analysis, of which 148 were rejected as a result of inadequate measures of retention. A final sample of 92 references was retained for use in a contents meta-analysis, while 42 references met the criteria for use in an effects meta-analysis.

Predictors Used in the Study of Retention in Higher Education

Background Characteristics

A list of predictors was generated based on conventional models of undergraduate student retention. Every study in the sample was coded as either having included each of the selected predictors or not (e.g., 0=not present, 1=present). Background characteristics included demographics, academic performance, and personality.

Constructs versus Measures

Each background characteristic listed in Table 1 represents a *construct*, which is an unobservable theoretical variable. Personality, for instance, is a theoretical construct used to explain behavior patterns in individuals. Manifest indicators, or *measures*, are observable variables that can be measured and are considered, based on theory, to reflect an underlying construct. Although we are unable to measure constructs directly, we may make inferences about them by measuring their characteristics. We will refer to these constructs as Level 1 Constructs, in that they are the constructs that were evidently being used in the primary research. This

will distinguish them from the Level 2 Constructs that were derived by us (the meta-analysts) from the patterns of association observed among the Level 1 Constructs. The Level 1 Constructs thus refer to the putative characteristics of the *students* being engaged as research participants in the primary research; the Level 2 Constructs, to be defined below, refer instead to pattern discerned among the characteristics of the *studies* that were sampled by our meta-analysis.

Table 1
Background Characteristics: Constructs and Measures

Level 1 construct	Measure
Demographics (Including SES)	Ethnicity
	Gender
	Father's level of education
	Mother's level of education
	Family income
	Parent occupation
	Household items
Past academic performance	ACT scores
	SAT scores
	Score on placement exams
	High school rank
Personality	Attitudes
	Self-concept
	Ethnicity
	Gender
	Father's level of education

Person-Role Fit

Person-Role Fit is another traditionally cited predictor of retention and includes factors such as academic and social integration, each of which is composed of several measures which are specified in Table 2.

Commitment

Commitment is the final category of conventional predictors of retention in higher education and is composed of goal commitment, institutional commitment, financial aid, and institutional characteristics, each with their own corresponding measures which are listed in Table 3.

CONTENTS META-ANALYSIS: ILLUSTRATION

Table 2

Person-Role Fit: Level 1 Constructs and Measures

Level 1 construct	Measure
Academic integration	Current grade point average Frequency of academic involvement - Use of library - Hours of study - Counseling/advising sessions Contacts with faculty/students regarding academic issues Development of work skills Development of intellectual growth Development of intellectual life Satisfaction with quality of instruction Satisfaction with quality of curriculum Perception of academic experience
Social integration	Attendance at social functions Attendance at student extracurricular activities Friends on campus Weekends spent on campus/month Informal contacts with faculty

Table 3

Commitment: Level 1 Constructs and Measures

Level 1 construct	Measure
Goal commitment	Commitment to earning degree Highest expected degree
Institutional commitment	Attitudes about institution Choice of current institution
Current financial status	Financial aid Scholarships Loans
Institutional characteristics	Size of institution (2 year, 4 year)

Non-Traditional Predictors

In addition to the factors typically investigated in studies of undergraduate retention, we generated a list of predictors that may be generally described as “student-centered” variables. These variables tap into the personal characteristics of students that may result in different program completion outcomes, such as the influence of significant others,

alienation, and academic culture. A complete list of the Level 1 Constructs and measures used is available in Table 4.

Table 4
Alternative “Student-Centered” Predictors

Level 1 construct	Measure
Influence of significant others	Encouragement or discouragement by parents, teachers, friends
Alienation	Opportunity cost (alternative life options Disparity between students’ expectations of involvement (academic/social) and actual involvement
Financial need	Contact with home Outside employment and number of hours Satisfaction with cost of attending How much of education paid by parents
Student culture	Emphasis of social elements of institution - Intercollegiate athletics - Academic, scholarly, intellectual - vocational, occupational Quality of relationships with other students, student groups, activities, faculty, administration, etc.
Institutional culture	Values Satisfaction with institutional prestige Helpfulness of university staff (Advisors/secretaries, bookstore/parking/cashiers)
Living on campus	Where students plan to live (on campus versus at home)
Prior college experiences	Attending college program prior to enrollment

Methodological Characteristics

Finally, two key methodological characteristics were included as meta-analytic predictors, and these are listed in Table 5. The first is Representativeness of Sample, which is coded as reported (1) or not (0). The second methodological characteristic is Reporting of Partial Weights (i.e., reporting unique variance of a predictor), again reported as present (1) or not (0).

Table 5
Methodological Characteristics of Studies

Level 1 construct	Measure
Representativeness of sample	Statistical comparison of study sample to student population
Reporting of partial weights	Statistical control of correlated effects

Contents Meta-Analysis of Predictors Used in the Study of Retention in Higher Education

Descriptive Statistics (Frequencies of Use)

As stated earlier, 92 studies were retained for contents analysis. For each of these studies, variables of interest were coded for the presence or absence of the variables listed above. Of course, only some studies reported variables on each construct. The frequencies of reports for each subordinate construct are listed in Table 6.

Empirical Relations among Predictors (Exploratory Factor Analysis)

The coded variables were analyzed in an exploratory factor analysis. Exploratory factor analysis is a method of identifying latent constructs by measuring shared variance between variables. However, because it is “exploratory”, the resulting latent constructs are simply mathematically constructed, rather than constructed using a priori justification. It is the responsibility of the investigator to evaluate each factor in terms of its theoretical value, and then to test their factor structure on an independent sample (i.e., confirmatory factor analysis).

One product of an exploratory factor analysis is the identification of patterns in variable measurement; in other words, we are able to see which variables tended to be measured together in the same study. These patterns have both conceptual and statistical value. Conceptually, we may begin to see the associations between variables which may reflect an underlying theoretical approach toward solving a particular problem. Statistically, in conducting a factor analysis, we are essentially attending to the problem of multicollinearity by reducing the number of variables in our data set. Failing to account for statistical dependence distorts our parameter estimates, but by condensing variance into fewer factors, we are reducing the degrees of freedom in our model and avoiding overspecification. As with any model specification, our objective is to maximize explanatory power with the fewest number of predictors.

Table 6
Percent of Studies Measuring Variable of Interest

	Percent
Background characteristics	
Demographics	78%
Past academic performance	64%
Personality	20%
Financial aid	24%
Current grades	25%
Person-role fit	
Academic integration	51%
Social integration	50%
Commitment	
Goal commitment	46%
Institutional commitment	26%
Alternative “student-centered” predictors	
Influence of significant others	23%
Living on campus	02%
Prior college experiences	01%
Financial need	10%
Methodological predictors	
Representativeness of sample	13%
Reporting of partial weights	27%

The factor pattern of each factor is illustrated in Table 7. These factor loadings were produced using Promax (oblique) factor rotation. Ideally, each construct loads on to one factor, and not on to other factors, so that each factor is a unique and independent structure. Conceptually, this factor pattern shows that the predictors that load on to a particular factor tend to be measured together.

We interpreted Factor 1 (the salient indicators being Academic Integration, Social Integration, Goal Commitment, and Institutional Commitment) as representing a mainstream theoretical approach in which dropping out is predicted by insufficient integration and affiliation (Bean, 1982; Tinto, 1975). Factor 2 (the salient indicators being the Influence of Significant Others, Financial Need, and Representativeness of the Sample) was interpreted as an alternative theoretical approach based on a student-centered model, where characteristics of the students and their personal life experiences become the focal point. Finally, Factor 3 (the salient indicators being Demographics, Past Academic Performance, Financial Aid, and Current Grade Point Average) was interpreted as the conventional empirical approach in which student interactions with an institution are emphasized.

Table 7
Rotated Factor Pattern (Standardized Regression Coefficients)

Level 1 constructs: Indicator variables	Level 2 constructs: "Research paradigm factors"		
	Factor 1 (Tinto-Bean models)	Factor 2 (Student- centered models)	Factor 3 (Institutional records models)
Demographics	-0.01	0.09	0.38
Past academic performance	-0.12	0.20	0.41
Personality	0.08	-0.26	-0.10
Financial need	0.15	0.09	0.63
Current GPA	0.06	-0.09	0.32
Academic integration	0.79	0.02	-0.05
Social integration	0.86	0.06	-0.02
Commitment to educational goals	0.68	0.07	0.04
Commitment to institution	0.78	-0.36	0.15
Influence of significant others	0.25	0.51	-0.43
Living on campus	-0.04	-0.00	0.03
Prior college experience	0.06	0.08	0.07
Financial need	0.16	0.52	-0.21
Representativeness of sample	-0.19	0.73	0.24
Partial correlation	0.31	0.49	0.43

To quantitatively test whether the factors were independent, we measured the correlations between them (see Table 8). We found that Factor 1 showed a small to moderate correlation with Factors 2 and 3. Factors 2 and 3 did not appear to be correlated.

Thus, the factor analysis produced three clear and distinct latent common factors that cumulatively explained 89% of the variance in the variables included in this analysis. The eigenvalues and proportions of variance explained for each factor, along with the cumulative proportion of variance explained, are outlined in Table 9. The optimal number of factors to retain were selected by means of a subjective scree test (Gorsuch, 1983),

Table 8
Factor Correlations

Level 2 constructs	“Research paradigm factors”		
	Tinto-Bean ¹	Student-centered ²	Institutional records ³
Factor 1 (Tinto-Bean models)	1.00		
Factor 2 (Student-centered models)	0.26	1.00	
Factor 3 (Institutional records models)	0.14	-0.04	1.00

Note: ¹Academic Integration, Social Integration, Goal Commitment, and Institutional Commitment; ²Influence of Significant Others, Financial Need, and Representativeness of the Sample; ³Demographics, Past Academic Performance, Financial Aid, and Current Grade Point Average.

as well as by the fact that adding more factors after accounting for 89% of the variance would likely be a waste of effort and model parsimony. We refer to these common factors as *Level 2 Constructs* because they do not directly describe the population(s) of *students* addressed by the primary researchers, but instead describe the population(s) of *studies* addressed by our meta-analysis. The Level 2 Constructs are, of course, based on the Level 1 Constructs that were originally measured in the primary research, but describe the investigative behavior of the primary researchers who elected to assess these particular constructs as opposed to all the possible alternative predictors that might have been imagined to influence student retention. The *students* are therefore the subjects of investigation for the primary researchers, whereas the *studies* performed by those researchers

Table 9
Eigenvalues of the Reduced Correlation Matrix

Model parameter	“Research paradigm factors”		
	Tinto-Bean ¹	Student-centered ²	Institutional records ³
Eigenvalue	3.05	1.34	1.23
Proportion of variance explained	0.48	0.21	0.20
Cumulative proportion of variance explained	0.48	0.70	0.89

Note: ¹Academic Integration, Social Integration, Goal Commitment, and Institutional Commitment; ²Influence of Significant Others, Financial Need, and Representativeness of the Sample; ³Demographics, Past Academic Performance, Financial Aid, and Current Grade Point Average.

are the subjects of investigation for the meta-analysts. These “Research Paradigm Factors” were therefore generated by the behavior of the researchers and not that of the original research participants.

Summary of Contents Meta-Analysis

Using exploratory factor analysis, we were able to identify three common factors corresponding to clusters of methods that tended to be used in conjunction with one another. These common factors appeared to reflect independent research paradigms or approaches to understanding student retention in higher education (see Table 10). Our next step will be to include these factors as predictors in the *effects* meta-analysis to determine whether one’s theoretical approach and one’s methods impact the magnitude of the mean effect sizes reported.

Table 10
Summary of Results of Contents Meta-Analysis

Level 2 constructs: Research paradigm factors	Level 1 constructs: Salient indicators
Mainstream theoretical approach (Tinto-Bean models)	Academic integration
Alternative theoretical approach (Student-centered models)	Social integration
	Goal commitment
	Institutional commitment
Conventional empirical approach (Institutional records models)	Influence of significant others
	Financial need
	Representativeness of sample
	Demographics
	Past academic performance
	Financial aid
	Current grade point average

Effects Meta-Analysis of Predictors Used in the Study of Retention in Higher Education

Now that we identified three unique research paradigms in undergraduate retention studies, we assess their influence on the effect sizes reported. We begin by reviewing the results of the synthesized effect sizes for each predictor. The subsequent step places the research paradigm factors derived in our previous analysis and the individual predictors into a General Linear Model in order to determine their

contribution to study outcomes. Finally, we demonstrate how hierarchical model comparisons can be used to test the unique contributions of factors, predictors, and interactions to reported effect sizes. The results of these comparisons serve as the foundation to create structural models and estimate the parameters of significant variables.

Descriptive Statistics (Effect Sizes of Predictors)

Table 11 shows the descriptive statistics for the synthesis of effect sizes of individual predictors. The means column is presented using standardized regression coefficients. The values represent the effect size between a given predictor variable and the outcome, retention. All predictors are positively related to retention, as expected. The lowest mean effect size ($r = 0.03$) is associated with Institutional Commitment, while the highest mean effect size ($r = 0.33$) reflects the contribution of Living on Campus.

Table 11
Descriptive Statistics for Effect Sizes (Standardized Regression Coefficients)

Variable	n	Mean	Min.	Max.
Demographics	29	0.17	0	0.44
Past academic performance	33	0.20	0	0.91
Personality	8	0.16	0	0.40
Financial aid	8	0.14	0	0.36
Academic integration	19	0.09	0	0.48
Social integration	19	0.07	0	0.46
Commitment to educational goals	20	0.16	0	0.90
Commitment to institution	5	0.03	0	0.25
Influence of significant others	11	0.05	0	0.21
Living on campus	2	0.33	0.29	0.36
Prior college experience	1	0.12	0.12	0.12
Financial need	3	0.10	0.07	0.37

Inferential Statistics (Meta-Analytic Predictive Models)

We now employ the predictors and factors in a hierarchical General Linear Model. The model is similar to a split-plot design, where sub-categories (“split-plots”) are nested within larger categories (“whole-plots”). Originally, Fisher (1925) proposed split-plot methods for agricultural experiments. Formal split-plot designs use independent treatment variables and randomly assign conditions at both the whole-plot and split-plot levels. For example, in a study with two independent variables (A and B), each with three levels (1-3), a split-plot design would

employ one independent variable as the whole-plot variable (e.g., A), and the other independent variable as the split-plot variable (e.g., B). Each level of the split-plot variable would be nested within each level of the whole-plot variable, producing nine different outcomes (A1 contains B1, B2, B3; A2 contains B1, B2, B3; A3 contains B1, B2, B3). It may also use larger blocking variables (superordinate to whole-plot categories) for replication purposes.

In the case of our model, we do have nested variables, where independent predictors are nested within the research paradigm factors. However, because factors were derived from predictors, we do not have a fully crossed design. Nevertheless, the analogy to “split-plot” design is useful in terms of understanding how error is treated. There are separate error terms for whole-plots and split-plots, so in the case of this example, we calculate the effect of paradigm factors and the associated residual error term, and the effect of predictors and factor-predictor interactions, and the associated residual error term (see Table 12).

We found that the main effects of research paradigm factors explained 23% of the variance in effect sizes in our sample. Combined with the associated error term, the between-studies variables amounted to a full 56% of the explained variance. Once paradigm factors were accounted for, the main effects of the predictor variables accounted for only 8% of the explained variance; however, factor-by-predictor interactions contributed another 13%. The entire model explained about 80% of the variance in the effect sizes of our sample.

Table 12
“Split-Plot” Hierarchical General Linear Model

Source of variance (heterogeneity)	Proportion of variance explained (sR^2)
Main effects of factors	.23
Residual effects of studies	.32
Subtotal between studies	.56
Main effects of predictors	.08
Factor*predictor interactions	.13
Subtotal within studies	.21
Whole model	.80
Residual	.23
Total	1.00

The General Linear Model provided us with main effects of factors and predictors. The next step of our analysis extends these findings by producing significance tests for the unique contributions of selected variables in our model. Using hierarchical modeling, it is possible to parse out the influence of a given variable and quantitatively test whether it

significantly adds to the overall model. These analyses are reported in Table 13.

Table 13
Hierarchical Significance Tests for Meta-Analytic Structural Hypotheses

Source	NDF	DDF	$F(H_0)$	$p > F$
Factor 1 (Tinto-Bean models)	1	38	26.93	0.0001
Factor 2 (Student-centered models)	1	38	0.14	0.7112
Factor 3 (Institutional records models)	1	38	0.09	0.7653
Study	38	157	2.92	0.0001
Predictor	11	157	2.60	0.0072
Tinto-Bean models Factor*predictor	9	157	1.50	0.1634
Student-centered models Factor*predictor	9	157	1.43	0.1888
Institutional records models factor*predictor	8	157	2.33	0.0264

Our model comparisons indicated that Factor 1 (Tinto-Bean models), Study, Predictor, and the Factor 3-by-Predictor Interaction (where Factor 3 is institutional records) were all significant. In other words, we can eliminate the remaining variables because they do not provide a significant contribution to our model of student retention in higher education.

We are now situated to describe the role of our variables in terms of parameter estimates. Parameter estimates offer value over and above standard null hypothesis significance testing (NHST), the limitations of which have been described extensively elsewhere (e.g., Cohen, 1994; Meehl, 1990). They are not subject to the assumptions of NHST, and offer an estimation of the parameters of the population of interest (Figueredo & Olderbak, 2008).

We have reduced our model to Factor 1 (Tinto-Bean Models), Study, Predictor, and Factor 3 (Institutional Records Models)-by-Predictor Interaction, and the parameter estimates for each of these variables are reproduced in Table 14. The parameter estimates for each variable have been tested for significance, resulting in further data reduction. Within this restricted model, the parameter estimates for Factor 1 (Tinto-Bean Models), Goal Commitment, and Factor 3 (Institutional Records Models)-by-Past Academic Performance are all significant. Demographics, Social Integration, and Factor 3 (Institutional Records Models)-by-Demographics are all moderately significant.

Table 14
Hierarchical Parameter Estimates for Statistically Significant Effects

Model parameter	Estimate	SE	$t(H_0)$	$p > t$
Intercept	0.19	0.01	14.67	0.0001
Factor 1 (Tinto-Bean models)	-3.12	0.46	-6.81	0.0001
Predictors				
Demographics	0.03	0.02	1.76	0.0812
Past academic performance	0.01	0.02	0.67	0.5032
Personality	-0.09	0.08	-1.11	0.2703
Financial aid	-0.03	0.09	-0.38	0.7047
Academic integration	-0.02	0.02	-0.67	0.5044
Social integration	-0.04	0.02	-1.74	0.0846
Commitment to educational goals	0.05	0.02	2.63	0.0096
Commitment to institution	-0.03	0.07	-0.37	0.7129
Influence of significant others	-0.07	0.07	-1.00	0.3214
Living on campus	-0.02	0.19	-0.13	0.8947
Prior college experience	-0.01	0.13	-0.04	0.9646
Financial need	0.32	0.24	1.31	0.1918
Factor 3 (Institutional records models)*predictor				
Demographics	-1.41	0.79	-1.78	0.0767
Past academic performance	2.64	0.74	3.58	0.0005
Personality	0.68	3.56	0.19	0.8490
Financial aid	-0.48	2.81	-0.17	0.8645
Academic integration	-0.32	0.89	-0.36	0.7179
Social integration	-0.37	0.78	-0.47	0.6384
Commitment to educational goals	-0.90	0.71	-1.28	0.2044
Commitment to institution	-1.07	2.20	-0.48	0.6294
Influence of significant others	-0.42	2.36	-0.18	0.8608
Living on campus	-0.75	5.86	-0.13	0.8984
Financial need	13.64	10.52	1.30	0.1969

The data in Table 14 indicate that studies that measure certain variables as predictors of undergraduate retention report different effect sizes. For example, studies which incorporated measurement variables associated with the Tinto-Bean Models of retention reported significantly lower effect sizes than the other models. Studies which examined retention by looking at institutional records, particularly past academic performance, reported significantly higher effect sizes than other conventional empirical variables. Non-significant predictors, in this case, do not statistically influence the magnitude of the effect size reported.

In essence, we are able to identify, quantitatively, how the methods used by the researchers in our sample impact the outcome reported. It does not, unfortunately, indicate the “true” effect size in the population of

students, but we may use these results to generalize to the population of research articles published in the scientific literature. We may also attempt to statistically account for inflated or deflated effect sizes in our estimates of the mean effect size.

To further investigate the relationship between methodology and reported outcomes, we calculated the mean effect size ($r=0.14$) in 42 studies that actually reported Squared Multiple Correlations.

Table 15 shows the results of a bivariate regression estimating the effect of the level of Factor 1 (Tinto-Bean Models) upon the magnitude of the Squared Multiple Correlations reported in the articles that actually reported the Squared Multiple Correlations. This means that a higher level of Factor 1 was predictive of a significantly lower Squared Multiple Correlation (the baseline being ~15% of the variance and the slope representing a loss of ~1% of the variance for every unit increment in Factor 1).

Table 15
Inferential Statistics: Paradigmatic Effects on R-Squared of Whole Model

Parameter	Estimate	SE	$t(H_o)$	$p > t$
Intercept	0.15	0.01	16.47	0.0001
Factor 1 (Tinto-Bean model)	-0.96	0.36	-2.67	0.0110

In the 60 articles that actually reported the Proportion of the Population Sampled, about 44% of the population was sampled, on average. Unfortunately, we do not know whether the participants sampled in these studies are representative of the population, but the percentage may provide an initial assessment of representativeness (such as if the percentage is very high).

Table 16 shows the results of a bivariate regression estimating the effect of the level of Factor 1 upon the Proportion of the Population Sampled reported in the articles that actually reported the Proportion of the Population Sampled. This means that a higher level of Factor 1 was predictive of a significantly higher Proportion of the Population Sampled (the baseline being ~44% of the population and the slope representing a gain of ~4.7% of the population for every unit increment in Factor 1).

Table 16
Inferential Statistics: Paradigmatic Effects on Proportion of Population Sampled

Parameter	Estimate	SE	$t(H_o)$	$p > t$
Intercept	0.44	0.03	14.63	0.0001
Factor 1 (Tinto-Bean model)	4.78	1.33	3.61	0.0006

A related observation is that, for the 92 studies used in the Contents Meta-Analysis, only 12 reported having collected a Representative Sample (regardless of the proportion of the population included), whereas 80 reported not having collected a Representative Sample of the population. The mean difference in the magnitude of the effect sizes between studies that did and did not report having collected a Representative Sample was .562, indicating that a more representative sample generally produced larger effect sizes.

Generalizability Theory Meta-Analysis

The Estimation of Variance Components

Generalizability Theory (GT) meta-analysis is another complementary method of meta-analysis. GT is based on the same theoretical variance-component model upon which all of the analysis of variance (ANOVA) is fundamentally based. In traditional ANOVA models, the hypothetical variance components presumably underlying the Observed Mean Squares (OMS), or observed variances, are used to identify the proper numerators and denominators needed for the construction of valid F-Ratios for the purpose of Null-Hypothesis Significance Testing (NHST). This is known as “Expected Mean Squares Analysis”, because the hypothetical variance components are often referred to as the Expected Mean Squares (EMS), as contrasted with the OMS. In GT models, the relative magnitudes of these hypothetical variance components are instead estimated for the purpose of constructing more direct quantitative comparisons among them, as opposed to the binary decisions of NHST.

The statistical theory underlying variance-component models specifies the hypothetical composition of the Observed Mean Squares (OMS) in terms of these constituent Expected Mean Squares (EMS). As with a chemical formula, which specifies not only what *kind* of atoms (classified as chemical “elements”) constitute a molecule (the basic unit of a chemical “compound”), but also their relative *numbers* in each molecule by means of an associated coefficient (typically as a subscript following each element), a complete EMS Analysis specifies which particular EMS components are hypothesized to comprise any given OMS in the empirical results, with an associated numerical coefficient (usually shown preceding each variance component) indicating the relative quantitative contribution of each to the total variance observed within each term.

Because these EMS expressions are formulae that are easily deduced from the basic design of any given study (even automatically by statistical software), the EMS Analysis is something that follows logically from the

table of observed mean squares and their associated degrees of freedom (which serve as the basis for deriving the coefficients assigned to each variance component). All that remains to pave the way for a true GT analysis is to use this information to estimate the expected magnitudes of the hypothesized variance components by solving the results of an EMS Analysis as a system of linear equations, with the OMS (or observed variances) and the numerical coefficients serving as the “known” quantities and the EMS (or expected variance components) serving as the “unknown” terms to be obtained through this procedure. This estimation may be performed by means of algorithms such as Ordinary Least Squares (OLS) or Restricted Maximum Likelihood (REML). The benefits of using REML include the fixing of the lower permissible bounds of variance components estimates to zero, whereas OLS sometimes produces negative estimates for variance components, which are generally interpreted as indicating zeros, but are technically out of proper bounds.

In our illustrative empirical example, Table 17 displays the estimated magnitudes of several of the major meta-analytic predictors identified in the hierarchical regression models presented above. These quantities are not very directly interpretable except in terms of their relative magnitudes. For example, we might note that the variance component estimated for the “random effect” of unexplained (residual) component of the variance in effect size among studies is over three times as large as that estimated for systematic variance attributable to the “focal” effect (meaning the effect of interest), or the level of Factor 1 associated with each study. However, this initial assessment is not the most pragmatically useful or interpretable result produced by a GT model. Those are given instead by what have been called *Generalizability Coefficients*.

Table 17
Variance Component Parameters for Generalizability Analysis

Variance component	Estimate
$\sigma^2(\text{Factor1})$	0.00006868
$\sigma^2(\text{Study})$	0.00022467
$\sigma^2(\text{Predictor})$	0.00005328
$\sigma^2(\text{Factor1*predictor})$	0.00004686

The Estimation of Generalizability Coefficients

Generalizability Theory (GT) is a direct extension of Classical Test Theory (CTT), which was originally proposed to address psychometric questions of reliability. According to CTT, any observation is composed of two components: “true score” variance and “error” variance. In this view,

all systematic variance is interpreted as an individual’s “true” score, or the value that would be obtained over infinite observations. All remaining variance is termed “error” variance, and is assumed to be composed only of measurement error.

In contrast, GT analysis disaggregates these variances further to account for components of the systematic variance attributable to variables other than the one under immediate study. These sources of variance, termed “facets”, can then be quantified so that their individual and interactive effects are modeled directly. A facet may be “fixed” or “random”, and these distinctions are analogous to “fixed effects” and “random effects” in the general linear models described above. In fixed effects models, the observed levels of a nominal variable (ANOVA “factor”) are considered to be exhaustive of the population of possible levels and generalizability is limited to the levels represented in the study. In random effects models, these observed levels are instead conceptualized as being randomly selected from a larger population; thus, generalizability beyond the current sampling design is presumed to be possible. In generalizability analyses, a fixed facet may produce systematic variance that can be modeled and predicted, while a random facet can be quantified to account for remaining heterogeneity.

To illustrate the difference between CTT and GT analysis, we will compare and contrast the estimation algorithms for the coefficients favored by each perspective. As stated earlier, the objective of CTT was to improve reliability in measurement. Its key metric is the reliability coefficient, which is the ratio between true score variance and error variance:

$$E^2_{\text{rel}} = \sigma^2_t / (\sigma^2_t + \sigma^2_e)$$

where σ^2_t is true score variance and σ^2_e is error variance. The generalizability coefficient is also a similarly-constructed “signal-to-noise” ratio, but it accounts additionally for multiple sources of “error” as well as their interactions with the fixed effects:

$$E^2_{\text{rel}*} = \sigma^2_f / (\sigma^2_f + \sigma^2_{\text{rel}*})$$

* if r is *nested* within f: $\sigma^2_{\text{rel}} = \sigma^2_{r(f)}$
 * if r is *crossed* with f: $\sigma^2_{\text{rel}} = \sigma^2_{r*f}$

By means of systematically applying these algorithms to the estimated variance components shown in Table 17, we are able to compute the GT Coefficients displayed in Table 18.

Table 18
Generalizability Theory Coefficients

Focal and random effects	GT coefficients
$\sigma^2(\text{Factor1})/(\sigma^2(\text{Factor1})+\sigma^2(\text{Study}))$.23
$\sigma^2(\text{Study})/(\sigma^2(\text{Study})+\sigma^2(\text{Study*Predictor}))$.25
$\sigma^2(\text{Predictor})/(\sigma^2(\text{Predictor})+\sigma^2(\text{Factor1*Predictor}))$.53
$\sigma^2(\text{Predictor})/(\sigma^2(\text{Predictor})+\sigma^2(\text{Study*Predictor}))$.07
$\sigma^2(\text{Factor1*Predictor})/(\sigma^2(\text{Factor1*Predictor})+\sigma^2(\text{Study*Predictor}))$.07

These numbers are substantially more interpretable than those for the raw variance components. For example, consistent with the ratio of more than three to one described above, the GT coefficient representing the “generalizability coefficient” of the influence of Factor 1 upon effect size magnitude over the remaining (“unexplained” or “residual”) variance among studies is .234, indicating that this systematic influence does not generalize very well across all the other “heterogeneous irrelevancies” or “local molar conditions” presumably underlying the residual variance among the effect sizes of the studies. This will facilitate the process of generalizing causal relations across studies (Sechrest, Parrin, & Bunker, 1990; Sechrest & Scott, 1993). Contrasting this result with those of the hierarchical regression model previously reported for these same data, in which the influence of Factor 1 was shown to be “statistically significant” at a quite respectable level of $p < 0.0001$ by NHST, indicating that the magnitude of this influence is almost definitely greater than zero, we see that this influence is rather weak in that it is mostly overwhelmed by the unexplained variability among effect sizes that remains.

The Relative Utility of GT versus Conventional Meta-Analytic Models

Both GT and conventional meta-analytic models represent special cases of variance component models within the general framework of hierarchical linear models (Bryk & Raudenbush, 1992). While conventional effects meta-analyses focus on structurally representing sources of heterogeneity within a sample of studies on an outcome, GT meta-analytic models allow us to quantify the degree of relative influence of any given predictor within a sample with respect to the other sources of variance.

This added function is an important one because a “highly significant” source of heterogeneity might discourage a meta-analyst from estimating a synthetic effect size for a metapopulation (see Figueredo, Black, & Scott, this issue), merely because the magnitude of that heterogeneity has been shown to be quantitatively greater than zero, and only “very probably” at

that. Thus, GT coefficients may provide the rationale for a researcher to aggregate moderately heterogeneous results while leaving open the possibility of reporting estimates of systematic variability that should be expected across studies. In short, we are not advocating for one method over another, but instead emphasizing that GT-based and more conventional meta-analytic models differ in their fundamental objectives, where the former are capable of extending beyond basic hypothesis-testing procedures to parameter-estimation methods aimed at quantifying the amount of systematic heterogeneity among effect sizes in a metapopulation, as well as parametrically assessing the relative strengths of the particular influences of any partial sources of this variance.

Summary and Conclusions

In this article, we presented an illustration of the implementation and interpretation of Contents Meta-Analysis using a sample of studies on retention in higher education. Through a sequence of steps involving exploratory factor analyses, hierarchical multiple regressions, and generalizability theory analyses, we were able to establish: 1) that it was possible to quantitatively identify patterns in the data that suggested the existence of three mutually discriminable research paradigms that reflected different theoretical approaches to studying retention in higher education; 2) a method for further data reduction by testing the explanatory power of these meta-analytic Level 2 Constructs with respect to the Level 1 Constructs sampled in the primary research; and 3) that there was a non-trivial quantitative impact of the methodological approaches captured by these Level 2 Constructs, as well as of the original Level 1 Constructs, on the reported outcomes (effect sizes) of the studies sampled. Then, we followed up by applying Generalizability Theory analysis to our findings to demonstrate the additional flexibility in data analysis and interpretation that researchers using meta-analysis may find useful for making decisions about aggregating heterogeneous data.

We have shown that three methodological approaches used in research on retention in higher education produce statistically different outcomes. These results have important meta-scientific implications. Model-driven science must be coupled with a rigorous comparison of research paradigms to make progress in science. Although this particular empirical example does not specify which model is “more correct”, Contents Meta-Analysis provides quantitative tools that may be used toward this end. For instance, one should ask *why* the Tinto-Bean model produces lower effect sizes than the Institutional Records or Student-Centered models. Can the decreased effect size be explained by the higher proportion of

population sampled? What other factors might contribute? Using this method, we gain insight into the characteristics of theoretical approaches and their associated methodological tools. Moreover, we may be able to assess how these differences affect our interpretations of data and our conclusions. Thus, this method has both model *testing* as well as model *building* applications.

There are, of course, limitations that require attention. As with any implementation of model-driven science, researchers must select relevant variables, taking care to avoid omission of those which are key to the investigation. Additionally, careful consideration of the relationships among constituent predictors in a model, both measured and unmeasured, is necessary to avoid distorted results. Similarly, meta-analysts are responsible for choosing appropriate study characteristics that potentially influence the relationships between predictors and outcomes. We have shown that, for many meta-analyses, simply coding basic methodological characteristics is not sufficient for answering meta-scientific questions. This conventional approach may function well for meta-analyses examining strictly controlled drug trials, but for the vast majority of meta-analytic projects, theory and empirical methods are variable across a sample of studies.

Additionally, there are practical considerations for widely used meta-analytic tools. Graphical representations of results, such as forest plots, typically show the relative strength of effect sizes across studies, but lack information about what might account for those differences. Using data from a Contents Meta-Analysis, we may supplement these conventional displays with forest plots representing mean effect sizes for a given research paradigm. Recent initiatives to standardize reporting procedures may also need to be amended to include additional methodological characteristics that would facilitate a Contents Meta-Analysis. A short, non-exhaustive list of candidate characteristics could include theoretical approach, variables and methods sampled and their relationships to theoretical approaches, discipline or field, application of technological advances, and “ecological” factors such as the types employed in this empirical example.

These methods hold great promise for use in various applied fields, such as Program Evaluation, where it is simply not a feasible option to avoid reporting an overall synthesis of the meta-analytic findings even when a finding of heterogeneity among effect sizes is obtained. A glib answer of “it depends” is simply an insufficient basis for the making of rational decisions by policymakers. If, on the other hand, one is able to provide a synthetic effect size for the meta-population, along with the associated generalizability coefficient across levels of any dimension of study design or analysis (at the level of the primary research) that is

evidently generating the observed heterogeneity, one can capture both the central tendency and the dispersion of this higher-level aggregate of trends.

References

- Bean, J. P. (1982). Conceptual models of student retention: How theory can help the institutional researcher. In E. Pascarella (Ed.) *New directions for institutional research: Studying student attrition*, 36, San Francisco: Jossey-Bass.
- Berger, J. B., Blanco Ramírez, G. & Lyon, S. (2012). Past to present: A historical examination of retention, p. 1-31 in A. Seidman (Ed.) *College Student Retention: Formula for Student Success*, (2nd ed.). Lanham, MD: ACE/Preager.
- Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models in social and behavioral research: Applications and data analysis methods first edition*. Newbury Park, CA: Sage Publications.
- Cohen, J. (1994), "The earth is round ($p < .05$)," *American Psychologist*, 49, 997-1003
- Figueredo, A. J., Black, C. J., and Scott, A. G. (2013). Complementary meta-analytic methods for the quantitative review of research. *Journal of Methods and Measurement in the Social Sciences*, 4, 21-45.
- Figueredo, A. J. & Olderbak, S. (2008). Commentary for special issue of journal of memory and language: Generalizability theory analysis for psycholinguistic applications. *Journal of Memory and Language*, 59, 426-433.
- Fisher, R. A. (1925). *Statistical methods for research workers*. London: Oliver and Boyd.
- Meehl, P. E. (1990). Why summaries of research on psychological theories are often uninterpretable. *Psychological Reports*, 66, 195-244.
- Pascarella, E. T. & Terenzini, P. T. (2005). *How college affects students: A third decade of research* (Vol. 2). San Francisco, CA: Jossey-Bass.
- Sechrest, L., Perrin, E., & Bunker, J. (Eds.). (1990). *Research methodology: Strengthening causal interpretations of nonexperimental data*. Rockville, MD: National Center for Health Services Research.
- Sechrest, L., & Scott, A. G (Eds.). (1993). *Special Issue: Understanding causes and generalizing about them, New directions for program evaluation*, 57. San Francisco: Jossey-Bass.
- Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research*, 45, 89-125.