Routledge
Taylor & Francis Group

TOOLS OF THE TRADE

Advantages of Integrative Data Analysis for Developmental Research

Sierra A. Bainter and Patrick J. Curran

University of North Carolina at Chapel Hill

Amid recent progress in cognitive development research, high-quality data resources are accumulating, and data sharing and secondary data analysis are becoming increasingly valuable tools. Integrative data analysis (IDA) is an exciting analytical framework that can enhance secondary data analysis in powerful ways. IDA pools item-level data across multiple studies to make inferences possible both within and across studies and can be used to test questions not possible in individual contributing studies. Some of the potential benefits of IDA include the ability to study longer developmental periods, examine how the measurement of key constructs changes over time, increase subject heterogeneity, and improve statistical power and capability to study rare behaviors. Our goal in this article is to provide a brief overview of the benefits and challenges of IDA in developmental research and to identify additional resources that provide more detailed discussions of this topic.

Recent work in developmental science has led to novel and complex theories aimed at understanding the development of memory, perception, cognition, problem solving, and language (e.g., Brune & Woodward, 2007; Gervain & Mehler, 2010; Hedrick, Haden, & Ornstein, 2009; Keen, 2011). Amid this progress in cognitive development research, high-quality data resources are accumulating, and data sharing and secondary data analysis are becoming increasingly valuable tools, particularly as grant funding becomes more competitive. Besides efficient use of competitive financial resources, data sharing promotes replication and integration of scientific findings, investigation of new hypotheses, and open scientific inquiry. Growing interest in secondary data analysis in developmental psychology is evident (e.g., Brooks-Gunn, Phelps, & Elder, 1991; Bullock, 2007; Friedman, 2007). Indeed, this very journal recently published an excellent summary of secondary data analysis and publicly available data sets relevant to research in cognitive development (Greenhoot & Dowsett, 2012). With the availability

---

Correspondence should be sent to Sierra A. Bainter, Department of Psychology, University of North Carolina, CB#3270 Davie Hall, Chapel Hill, NC 27599, USA. E-mail: sbainter@email.unc.edu

of multiple high-quality data sets, novel methodological and analytical tools are needed to take full advantage of these data.

One such novel methodological framework is integrative data analysis (IDA; Bauer & Hussong, 2009; Curran & Hussong, 2009; Hussong, Curran, & Bauer, 2013). Briefly, IDA is a promising set of methodologies that might be useful for enhancing secondary data analysis in powerful ways and facilitating synthesis in cognitive development research. IDA pools item-level data across multiple studies to make inferences possible both within and across studies. Depending on the characteristics of the contributing data sets, there are many potential advantages to using IDA over and above the secondary data analysis of any single contributing data set. Especially relevant to developmental research, IDA can be used to study longer developmental periods, examine how the measurement of key constructs changes over time, increase subject heterogeneity, and improve statistical power and capability to study rare behaviors. Ultimately, IDA is useful not only to support empirical replication, but also to test questions not possible in individual contributing studies. However, IDA also presents a unique set of challenges that are not typically salient when analyzing data from a singly study.

Current trends in science suggest that now is an important time for pooled data analysis efforts such as IDA. Funding agencies have introduced policies to encourage data sharing (e.g., National Institutes of Health [NIH], 2003), and the technology to store and distribute valuable data resources has advanced tremendously in recent decades. The NIH also fosters data sharing by supporting the development of high-quality, standard measures for researchers conducting diverse empirical studies in the behavioral sciences. Currently funded efforts include the NIH Toolbox, the Patient-Reported Outcomes Measurement Information System, and the PhenX toolkit. IDA can be used to meet many of the challenges posed by pooled data analysis and can allow researchers to capitalize on secondary data resources.

Our goal in this article is to provide a brief overview of what IDA is and how it can be used to advance developmental research. We will highlight the core challenges to conducting IDA, identify situations where it may or may not be useful, and direct interested researchers toward resources with more in-depth information about conducting IDA. We also hope to describe how a pooled analysis using IDA can be greater than the sum of its parts, and perhaps our modest contribution will spark creative ideas for IDA in cognitive development research.

DEFINITION OF IDA

IDA is an analytical framework used to pool raw data from two or more studies for combined analysis. The strategy of IDA is to use psychometric modeling techniques to link the measurement across studies and create *commensurate measures*, meaning measures with the same meaning and scale across studies despite differences in assessment instruments and modalities. IDA is not one standardized technique; rather, it is a guiding framework for combining raw data from multiple studies. Although many of the individual components underlying IDA are not novel, together, the IDA framework is an innovative way to take full advantage of secondary data resources. Pooling raw data from multiple studies can lead to added advantages, as well as challenges, compared with a single-study analysis or even separate analyses using data from multiple studies.

IDA is distinct from existing strategies to combine information across studies, such as meta-analysis. Whereas meta-analysis is used to analyze published summary statistics from a

large number of studies (Cooper, Hedges, & Valentine, 2009), the unique advantages and challenges of IDA stem from pooling at the level of the raw data. Data integration can be visualized on a continuum from analysis of single-study raw data at one extreme (least integrative) to combining summary statistics from many studies using meta-analysis at the other extreme (most integrative); IDA lies in between these two extremes (Curran & Hussong, 2009). Both of these approaches play important roles within any area of research, and we believe IDA is an important tool to add to our field's analytical toolbox. Now we will consider some of the potential advantages of secondary data analysis using IDA, followed by a description of some of the core challenges.

ADVANTAGES OF IDA

There are many potential benefits to performing IDA, all of which vary depending on the characteristics of the contributing data sets and the motivating hypotheses. Some of these benefits are simply related to a larger sample size: Greater power is achieved by merging multiple studies. The larger sample size from pooling studies also improves stability to study rare behaviors. Other advantages of IDA help meet the need in psychology to integrate and replicate findings. Linking studies through IDA provides a deeper understanding of how constructs develop over time, built-in replication in heterogeneous samples, and broader measurement of intended constructs.

Improve Understanding of Development

IDA can aid the understanding of development in two key ways. First, joining studies with overlapping ages allows for examination across substantially longer developmental periods than observed in any single contributing study. One study might follow children from ages 3 through 9 years, a second from ages 7 through 13 years, and a third from ages 10 to 18 years; this could allow for the pooled IDA estimation of development processes spanning ages 3 through 18 years old. For example, in our own work, we have used IDA to fit developmental models spanning ages 10 through 33 years old, yet no individual participant contributed more than five repeated measures (Curran et al., 2008). Joining studies to observe a broader swath of development is one key way that IDA can permit tests of hypotheses that cannot be tested within a single contributing study.

Second, the measurement approach we use in IDA can be used to measure theoretical constructs over time while accounting for heterotypic continuity (i.e., changing manifestations of the same underlying developmental process). Essentially, we are able to develop models using IDA that create scores for an intended construct that are on the same scale not only across studies, but also across ages or other important covariates. For example, our research group created IDA scores for internalizing symptoms using 13 items assessing anxiety and depression (Hussong, Flora, Curran, Chassin, & Zucker, 2008). Some items, such as ''cries a lot,'' are less strongly related to the underlying level of internalizing at younger ages, and we were able to statistically control for this heterotypic continuity. In cognitive development research, this ability to describe or control for processes of heterotypic continuity alone could be a fruitful focus of IDA applications.

Increase Sample Heterogeneity

IDA can also be used to increase sample heterogeneity. Many individual samples in developmental psychology underrepresent important racial, socioeconomic-status, or gender subgroups, and IDA can be used to pool studies that differ on these key characteristics. Pooling studies to create a larger, more diverse sample permits testing of hypotheses simultaneously in these groups. Likewise, IDA may permit comparisons across subgroups that cannot be compared in the individual studies due to small sample sizes. Similar results across heterogeneous studies bolster the external validity of findings. Conversely, discrepant findings across studies provide valuable clues as to why the discrepancy exists to generate new hypotheses. Whereas in a meta-analysis or literature review, it is only possible to speculate as to the cause of discrepant findings, IDA can be used to test and understand study differences.

Improve Measurement

When the joined studies measure the same construct in somewhat different ways (with some overlap; see requirements for IDA in the next section) the pooled measure will be more comprehensive than the measures used in either study. For example, if IDA is used to pool data from two studies that assess expressive language using different instruments (e.g., the Reynell Development Language Scales-Revised, Reynell & Huntley, 1985; and the Expressive Vocabulary Test, Williams, 1997), the resulting pooled measure will give a more complete assessment of expressive language than that used in either study alone. Essentially, IDA borrows the measurement strengths from each study and creates more informative scores for each target construct.

Not all of these advantages will be achieved in every application of IDA, and IDA will not be appropriate for all multistudy analysis problems. In the next section, we describe the core challenges that arise when conducting IDA and describe situations when IDA may not be feasible.

CORE STEPS IN INTEGRATIVE DATA ANALYSIS

Develop Novel Hypotheses

The first step toward successful IDA is to identify a multistudy theoretical question of interest. Considering the strengths of IDA for developmental research, IDA applications may be motivated by questions that call for studying an extended period of development or a larger sample size, or for examining relatively rare behaviors. IDA is also ideally suited for characterizing heterotypic continuity by identifying how the manifestations of the same underlying process change over time.

Identify Contributing Data Sets

For contributing data sets to be linked in IDA, each construct key to the theoretical question of interest must be assessed in each study. For example, if the motivating hypothesis of the IDA application concerns expressive language and knowledge acquisition, some items relevant to

both constructs are needed in each study. Existing databases relevant to developmental research are certainly valuable resources for IDA endeavors (see Greenhoot & Dowsett, 2012). However, applications for IDA are not limited to public-access data sets. In addition to established public-access databases, collaborative efforts among investigators are being encouraged by funding institutes (NIH, 2003). Direct collaborations among researchers for the purpose of IDA could lead to many exciting opportunities. In our own work, we have collaborated with investigators on three landmark longitudinal studies that focus on children of alcoholic parents (Chassin, Rogosch, & Barrera, 1991; Sher, Walitzer, Wood, & Brent, 1991; Zucker et al., 2000). Our pooled sample spans from childhood through adolescence and into adulthood (Curran et al., 2008). IDA can be used to pool existing data sets from completed studies as in our own examples, and furthermore, IDA also offers exciting opportunities for ongoing or future studies. Data collection efforts can be coordinated (e.g., measurement, sample characteristics) to facilitate IDA with existing data sets or other planned projects.

Assess Heterogeneity

It might seem that the ideal scenario for IDA would be one in which all studies have used identical ''gold standard'' measures for each construct, presenting all of the same items in an identical experimental design. Such a scenario is unlikely, and we believe working with studies that have assessed theoretical constructs of interest with a combination of similar and dissimilar items is actually beneficial. Although this more realistic scenario may be more challenging, performing IDA on studies with some heterogeneity in measurement, samples, measures, and experimental paradigms goes beyond exact replication and helps integrate findings within developmental science.

When determining the amount of between- and within-study heterogeneity, potential sources to consider include the sampling/selection procedures used in each study, ages assessed, different geographic regions, and study design characteristics. Besides the chronological ages of participants, historical period and birth cohort may be important factors (see Curran & Hussong, 2009, for more on between- and within-study sample heterogeneity). Some of these factors will be confounded by study; for example, if each study occurs in a different geographic region, it would not be possible to know if different findings between studies are due to geography or some other study characteristic. Although we cannot narrow in on the source of the discrepancy in such a case, we may still be able to control for these differences when creating commensurate measures. The amount of heterogeneity between studies cannot be excessive. For example, it is important to have sufficient overlapping ages to not completely confound study and age differences. Without this overlap, it is impossible to disentangle age and study differences.

Develop Item Pool

To link studies for IDA, *common items* are needed for each construct to link the measurement between studies. We define common items to mean items that have the same prompt and response scale. In developmental research, items can take many different forms, but some examples include responses to individual questions in a test or battery or endorsement of symptoms. It is not necessary for these items to be common to all studies; it is sufficient for common items to

link pairs of studies as long as there is enough overlap to link measurement across studies. Importantly, we can use items that are unique to one study; these unique items improve the precision of our measurement, but they do not help us link scores across studies (Curran & Hussong, 2009).

Often, the original response scale or prompt is not exactly the same in each study, but we can *harmonize* the measures by collapsing categories, combining items, and/or binning responses to arrive at a common item. In Table 1, a harmonization example drawn from Curran et al. (2008) shows how three studies used slightly different prompts and response scales to assess feeling fearful or anxious. This item is harmonized to a common dichotomous response scale of absent (0) or present (1).

Harmonization is an essential step in the IDA process, yet it is important to understand that it is not sufficient to produce commensurate measures across studies. This is because other differences, such as mode of administration, an item's context in terms of surrounding items, and how participants view their research participation in a particular location or context, can influence how participants in one study respond to an item, separate from actual levels of the underlying trait of interest. In the next section, we will describe how the psychometric approaches we use in IDA control for these study differences by testing and accounting for differential item functioning (DIF) and creating scores on the same scale in each study.

Develop a Measurement Model and Control for Study Differences

The next challenge in IDA is to develop a measurement model that can be used to create scores for subsequent analyses of the pooled data. The overarching aim for this process is to measure the intended construct while controlling for differences across important covariates. The chief concerns throughout this process are to ensure that models are defined properly within each study, are appropriate for the item set, and characterize the same construct in each study. From start to end, IDA centers on issues of sound measurement, which is clearly a central concern in the study of cognitive development.

A measurement model that is well suited for IDA is an extension of factor analysis and item-response theory (IRT) models and is referred to as moderated nonlinear factor analysis

TABLE 1
Example of Harmonizing Response Scales Across Three Studies

	Study 1	*Study 2*	*Study 3*	*Harmonized Item*
Prompt	I am too fearful or anxious.	I was too fearful or anxious.	Feeling fearful	Too fearful or anxious
Response Scale	0. Not true	0. Almost never	0. Not at all	0. Absent
	1. Sometimes or sometimes true	**1. Once in a while**	**1. A little bit**	**1. Present**
	2. Very often true	**2. Sometimes**	**2. Moderately**	
		3. Often	**3. Quite a bit**	
		4. Almost always	**4. Extremely**	

Note. Original values of ''0'' were harmonized to a value of ''0''; original values greater than ''0'' were harmonized to a value of ''1.''

(MNLFA; Bauer & Hussong, 2009). MNLFA is a confirmatory factor analytic model that allows the model parameters (e.g., factor mean/variance, item intercepts, factor loadings) to vary as a function of observed moderator variables (e.g., study, age) and also allows for nonlinear relationships between the latent factor and the indicators (e.g., binary or ordinal indicators). The individual items from the pool are used as indicators in the measurement model. Just as in factor analysis and IRT, this measurement model assumes that levels of the trait can be measured using an underlying continuum (latent factor). By this we mean that participants vary quantitatively in their level of the trait, but not qualitatively in their patterns of performance across items. A latent factor model assumes that as the underlying levels on the latent trait increase, the probabilities of scoring higher on each item simultaneously increase. We will briefly consider the two defining aspects of the MNLFA model: nonlinear relationships and moderator variables.

Importantly, the ''nonlinear'' relationships allowed between the latent factor and the items in MNLFA mean the model is not restricted to any one response distribution for the items. Whereas continuous indicators are assumed to be linearly related to the latent factor, different response distributions are needed to include binary, categorical, and count items. This is often essential in IDA applications, where response scales for items are mixed across and even within studies. MNLFA can allow a mix of continuous, categorical, binary, and count items, any of which can be unique to one study or common across studies (Bauer & Hussong, 2009). Appropriate response distributions can be specified for each indicator (e.g., logistic for binary items to bound the probability of endorsing an item between 0 and 1).

Another crucial component of the model-building process for IDA is to model differences in the latent factor mean and variance as a function of study, age, and other covariates. These effects are the ''moderated'' aspect of MNLFA, and they allow us to account for overall differences in mean level or variability in each study. Similarly, age effects incorporate growth trends, and gender effects can model a higher level for boys or girls. Including the effects of these and other important predictors improves the validity of the model and preserves useful variability to generate more informative scores. Our recommended model-building approach is to start with simple models and gradually build to more complex models as needed, with the goal to build the best, parsimonious model that accurately captures important variability in the factor mean and variances and controls for differences in the trait (i.e., mean and variance) by study and other important covariates.

After developing a general psychometric model for the factor mean and variance, the final model building step in IDA is to determine if any individual items do not behave identically within each study, across ages, and across any other important groups in the sample. To assess this, we test for DIF in each item. Adequately testing and accounting for potential DIF gives us confidence that the measurement is linked across studies and the resulting scores will be commensurate across the groups for which we have controlled. Testing for DIF by study allows us to statistically correct for the many differences in context or assessment methods that may influence participants to respond to an individual item differently in each study. For example, a slight variation in an item prompt may make participants more likely to strongly endorse an item in a particular study, and these subtle differences are not controlled by harmonizing items to a common response scale. Similarly, if an item (such as *easily cries*) is more highly endorsed and normative at younger ages, DIF by age is used to explicitly account for this heterotypic continuity. We use a sequential procedure to evaluate DIF item by item, similar to procedures used to evaluate measurement invariance in factor analysis (Yoon & Millsap, 2007) and the direct IRT method

for evaluating DIF (Embretson & Reise, 2000, pp. 252–262; Thissen, Steinberg, & Wainer, 1993). More details on DIF analysis in IDA and a review of the methods developed in the IRT and factor analysis traditions to evaluate whether an item is functioning identically for different portions of the sample are described in Bauer and Hussong (2009). After evaluating possible DIF, the model is ready for scoring.

Scoring and Hypothesis Testing

Scores can be generated for each observation using the developed psychometric model. Assuming assumptions hold, the obtained scores are commensurate across studies, age, and any other important predictors controlled for in model development. The final scores obtained will be continuous, on a standardized scale, and can be used for subsequent analyses with the pooled data to evaluate hypotheses. Whatever statistical model is used for subsequent hypothesis testing, it is still important to account for sources of between-study heterogeneity. This is usually most easily done by including the effects of study membership directly in each model. In our own work, we have most often used IDA scores to test hypotheses using multilevel models (Hussong, Cai, et al., 2008) and latent growth curve models (Hussong, Flora, et al., 2008). Hussong et al. (2013) provide more details on hypothesis testing in IDA.

WHEN IS IDA UNFEASIBLE?

IDA will not be appropriate for every multistudy application, yet there are few definitive rules as to when IDA becomes unfeasible. For instance, with insufficient overlap of items, IDA may not be possible or justifiable. Even if overlap between studies exists, the between-study differences may be too extensive for the application to be justifiable. For example, too few invariant items would decrease our confidence that the measurement is linked across studies.

Multiple studies may also introduce core confounds. If the study completely confounds a central question—for example, study and race are completely confounded—then it will be impossible to disentangle study and race differences. In this situation, if the driving theoretical question is primarily concerned with racial differences, IDA may not provide the necessary information. Complete lack of uniformity across studies would be a warning that IDA may not be possible for this particular item pool and set of studies.

In cases where studies are not suitable for IDA (due to insufficient overlap in items, ages, etc.), one promising option is to plan a new primary data collection to facilitate the IDA. We refer to this option as a bridging study. In the case of insufficient common items, a bridging study would involve administering common items in a new sample to help ensure that measurement can be linked across studies. Similarly, a bridging study could be designed to create overlap in ages or other important groups. Hussong et al. (2013) provide more details on planning a bridging study.

NEXT STEPS AND CONCLUSION

We hope that we have helped provide an initial understanding of the benefits IDA has to offer for developmental research. If you are interested in conducting an IDA study, a good next step

would be to consult empirical examples and pedagogically oriented articles that can help you walk through different aspects of the process. Hussong et al. (2013) provide an in-depth, nontechnical tutorial with examples and many details that are omitted here. Curran and Hussong (2009) and Bauer and Hussong (2009) provide more detailed overviews of IDA as well as more technical details related to MNLFA and scoring. Other helpful resources include Hofer and Piccinin (2009), who describe the design and execution of a coordinated analysis approach for pooling data resources. We also refer interested readers to detailed examples of analyses done using IDA including McArdle, Grimm, Hamagami, Bowles, and Meredith (2009), Curran et al. (2008), Hussong, Cai, et al. (2008), and Lorenz et al. (1997).

Many methodological challenges arise in this new and exciting vein of research, and advanced statistical training may be needed to overcome these challenges. The specific expertise needed will depend on the research question at hand. To help meet a growing need for advanced statistical training in psychology, a number of advanced statistical workshops are being offered across the country through universities and research institutes and at preconference workshops. We believe the potential benefits of IDA far outweigh the associated challenges, and this innovative research tool can enhance secondary data analysis in powerful ways to advance developmental research.

REFERENCES

Bauer, D. J., & Hussong, A. M. (2009). Psychometric approaches for developing commensurate measures across independent studies: Traditional and new models. *Psychological Methods*, *14*, 101–125. doi:10.1037/a0015583

Brooks-Gunn, J., Phelps, E., & Elder, G. H. (1991). Studying lives through time: Secondary data analyses in developmental psychology. *Developmental Psychology*, *27*, 899–910. doi:10.1037/0012-1649.27.6.899

Brune, C. W., & Woodward, A. L. (2007). Social cognition and social responsiveness in 10-month-old infants. *Journal of Cognition and Development*, *8*, 133–158. doi:10.1080/15248370701202331

Bullock, M. (2007). Secondary analysis—extending the value of data. *Journal of Applied Developmental Psychology*, *28*, 383. doi:10.1016/j.appdev.2007.07.002

Chassin, L., Rogosch, F., & Barrera, M. (1991). Substance use and symptomatology among adolescent children of alcoholics. *Journal of Abnormal Psychology*, *100*, 449–463. doi:10.1037/0021-843X.100.4.449

Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.). (2009). *The handbook of research synthesis and meta-analysis*. New York, NY: Sage.

Curran, P. J., & Hussong, A. M. (2009). Integrative data analysis: The simultaneous analysis of multiple data sets. *Psychological Methods*, *14*, 81–100. doi:10.1037/a0015914

Curran, P. J., Hussong, A. M., Cai, L., Huang, W., Chassin, L., Sher, K. J., & Zucker, R. A. (2008). Pooling data from multiple longitudinal studies: The role of item response theory in integrative data analysis. *Developmental Psychology*, *44*, 365–380. doi:10.1037/0012-1649.44.2.365

Embretson, S. E., & Reise, S. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.

Friedman, S. L. (2007). Finding treasure: Data sharing and secondary data analysis in developmental science. *Journal of Applied Developmental Psychology*, *28*, 384–389. doi:10.1016/j.appdev.2007.07.001

Gervain, J., & Mehler, J. (2010). Speech perception and language acquisition in the first year of life. *Annual Review of Psychology*, *61*, 191–218. doi:10.1146/annurev.psych.093008.100408

Greenhoot, A., & Dowsett, C. (2012). Secondary data analysis: An important tool for addressing developmental questions. *Journal of Cognition and Development*, *13*, 2–18. doi:10.1080/15248372.2012.646613

Hedrick, A. M., Haden, C. A., & Ornstein, P. A. (2009). Elaborative talk during and after an event: Conversational style influences children's memory reports. *Journal of Cognition and Development*, *10*, 188–209. doi:10.1080/15248370903155841

Hofer, S. M., & Piccinin, A. M. (2009). Integrative data analysis through coordination of measurement and analysis protocol across independent longitudinal studies. *Psychological Methods*, *14*, 150–164. doi:10.1037/a0015566

Hussong, A. M., Cai, L., Curran, P. J., Flora, D. B., Chassin, L. A., & Zucker, R. A. (2008). Disaggregating the distal, proximal, and time-varying effects of parent alcoholism on children's internalizing symptoms. *Journal of Abnormal Child Psychology*, *36*, 335–346. doi:10.1007/s10802-007-9181-9

Hussong, A. M., Curran, P. J., & Bauer, D. J. (2013). Integrative data analysis in clinical psychology research. *Annual Review of Clinical Psychology*, *9*, 61–89. doi:10.1146/annurev-clinpsy-050212-185522

Hussong, A. M., Flora, D. B., Curran, P. J., Chassin, L. A., & Zucker, R. A. (2008). Defining risk heterogeneity for internalizing symptoms among children of alcoholic parents. *Development and Psychopathology*, *20*, 165–193. doi:10.1017/S0954579408000084

Keen, R. (2011). The development of problem solving in young children: A critical cognitive skill. *Annual Review of Psychology*, *62*, 1–21. doi:10.1146/annurev.psych.031809.130730

Lorenz, F. O., Simons, R. L., Conger, R. D., Elder, G. H., Johnson, C., & Chao, W. (1997). Married and recently divorced mothers' stressful events and distress: Tracing change across time. *Journal of Marriage and Family*, *59*, 219–232.

McArdle, J. J., Grimm, K. J., Hamagami, F., Bowles, R. P., & Meredith, W. (2009). Modeling life-span growth curves of cognition using longitudinal data with multiple samples and changing scales of measurement. *Psychological Methods*, *14*, 126–149. doi:10.1037/a0015857

National Institutes of Health. (2003, March 5). *NIH data sharing policy and implementation guidance*. Retrieved from http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm

Reynell, J. K., & Huntley, M. (1985). *Reynell Developmental Language Scales-Revised, Edition 2*. Windsor, UK: NFER.

Sher, K. J., Walitzer, K. S., Wood, P. K., & Brent, E. E. (1991). Characteristics of children of alcoholics: Putative risk factors, substance use and abuse, and psychopathology. *Journal of Abnormal Psychology*, *100*, 427–448. doi:10.1037/0021-843X.100.4.427

Thissen, D., Steinberg, L., & Wainer, H. (1993). Detection of differential item functioning using the parameters of item response models. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 67–113). Hillsdale, NJ: Erlbaum.

Williams, K. (1997). *Expressive Vocabulary Test*. Circle Pines, MN: American Guidance Service.

Yoon, M., & Millsap, R. E. (2007). Detecting violations of factorial invariance using data-based specification searches: A Monte Carlo study. *Structural Equation Modeling*, *14*, 435–463.

Zucker, R. A., Fitzgerald, H., Refior, S., Puttler, L., Pallas, D., & Ellis, D. (2000). The clinical and social ecology of childhood for children of alcoholics: Description of a study and implications for a differentiated social policy. In H. Fitzgerald, B. Lester & B. Zuckerman (Eds.), *Children of addiction: Research, health, and policy issues* (pp. 109–141). New York, NY: Routledge Falmer.