Published in final edited form as:

Psychol Methods. 2009 June ; 14(2): 77–80. doi:10.1037/a0015972.

The Seemingly Quixotic Pursuit of a Cumulative Psychological Science: Introduction to the Special Issue

Patrick J. Curran

University of North Carolina at Chapel Hill

Abstract

The following manuscript is the final accepted manuscript. It has not been subjected to the final copyediting, fact-checking, and proofreading required for formal publication. It is not the definitive, publisher-authenticated version. The American Psychological Association and its Council of Editors disclaim any responsibility or liabilities for errors or omissions of this manuscript version, any version derived from this manuscript by NIH, or other third parties. The published version is available at www.apa.org/journals/met.

The goal of any empirical science is to pursue the construction of a cumulative base of knowledge upon which the future of the science may be built. However, there is mixed evidence that the science of psychology can accurately be characterized by such a cumulative progression. Indeed, some argue that the development of a truly cumulative psychological science is not possible using the current paradigms of hypothesis testing in single-study designs. The author explores this controversy as a framework to introduce the six papers that make up this special issue that is focused on the integration of data and empirical findings across multiple studies. The author proposes that the methods and techniques described in this set of papers can significantly propel us forward in our ongoing quest to build a cumulative psychological science.

The novel *El Ingenioso Hidalgo Don Quijote de la Mancha* by Miguel de Cervantes tells the story of the relationship between the delusionally idealistic Don Quixote and his world-weary yet practical companion Sancho Panza. Their adventures juxtapose the eager pursuit of unrealistic ideals with the practical grounding in the reality of day-to-day life. The main character succeeded in capturing the very nature of idealistic pursuit to the point that over time his name developed into an adjective that describes something as foolishly impractical. So is it fair to describe the pursuit of a cumulative psychological science as quixotic? On some days I believe that it is, while on others I do not.

Without question we as a field must in some way or another be motivated by a sincere idealistic quest to systematically build a cumulative base of knowledge upon which to progress the science of psychology. We simply cannot survive as a viable science in the absence of these ideals. Yet at the same time the vicissitudes of real life intervene in many forms including the pressures of publication, grant acquisition and tenure review; we must make practical decisions that may serve us on an individual level yet may not necessarily move our science forward in a cumulative fashion.¹ We thus find ourselves in a position where we need to each behave in a way that serves our own interests while at the same time contributing to the construction of a more cohesive and cumulative psychological science. In thinking how this might be best

Address for Correspondence: Patrick J. Curran, Department of Psychology, University of North Carolina, Chapel Hill, NC, 27599-3270, phone: 919.962.5235, fax: 919.962.2537, curran@unc.edu.

¹See Mischel (2008, 2009) for a wonderful exploration of these same issues.

accomplished, it is helpful to first consider how these issues play out in the general endeavor of scientific inquiry. An excellent starting point is to consider the work of Thomas Kuhn.

Arguably one of the most impactful works on the philosophy of science in the past century is Kuhn's *The Structure of Scientific Revolutions* first published in 1962 and revised in 1970 and again in 1996. Kuhn was writing in response to the dominant view of *logical empiricism* that was primarily held in the first half of the 20th century. This perspective held that science was a highly objective and logical process that led to the systematic construction of a cumulative understanding about the world around us (e.g., Howard, 1991). This was consistent with the notion that contemporary scientists were “standing on the shoulders of giants” in that the generation of new knowledge was systematically constructed upon the foundation of prior discoveries.

Within this dominant context, Kuhn offered the provocative (and what was to become highly controversial) perspective that science is indeed cumulative, but only during “normal” periods of relative calm where existing paradigms allow for the creation of knowledge to progress in a systematic fashion. However, as scientific anomalies for which existing paradigms cannot account begin to accrue, tension builds. Existing paradigms attempt to suppress these anomalies, but at some point the tension becomes too great and there is a “paradigm shift” in which the old paradigms are not simply augmented by the new ones, but instead are outright replaced. Kuhn considered these paradigm shifts to be scientific “revolutions” and uses examples such as moving from an earth-centered to sun-centered understanding of our solar system. Even nearly half of a century after the publication of his original work, skirmishes still break out over whether Kuhn's work implies that the very nature of science is or is not a cumulative process (see, e.g., D'Espagnat, 2008, and Weber, 2008, for recent competing perspectives) or if Kuhn really set out to kill logical empiricism in the first place (Reisch, 1991).

In the presence of the broader issues underlying the progression of the very endeavor of science itself, we must remain diligently focused on the more practical issues of how we can best support and build a cumulative science in psychology. Interestingly, there is as much disagreement as to whether we have achieved (or even can achieve) a cumulative science within psychology as there is to whether the very process of science itself is a cumulative endeavor. And, even more interestingly, some of our harshest critics quite appropriately reside among us. For example, in commenting on the state of psychological inquiry, Meehl (1978) remarked that “it is simply a sad fact that in soft psychology theories rise and decline, come and go, more as a function of baffled boredom than anything else; and the enterprise shows a disturbing absence of that *cumulative* character that is so impressive in disciplines like astronomy, molecular biology, and genetics” (p. 807; italics original). More recently, Schmidt (1996) stated that the reliance on significance tests in psychological research “...has systematically retarded the growth of cumulative knowledge in psychology” (p. 115). And Schmidt and Hunter (1997) recommend that researchers “...try to show enough intellectual courage and honesty to reject the use of significance tests despite the pressures of social convention to the contrary. Scientists must have integrity” (p.61).

Thus the dominant issue thought to be responsible for the failure of psychology to progress in a cumulative fashion is the over-reliance (or reliance in any fashion) on significance testing in single-sample analysis. Cohen (1990, 1994), Hunter and Schmidt (1996), and Schmidt (1996) all provide excellent discussions of the inherent limitations of this strategy. Yet single-study hypothesis tests remain our dominant paradigm: some statistical model is fitted to a single sample of data obtained through a single study and a go/no-go decision is made about a given effect primarily based on the obtained *p*-value. One widely recommended (although arguably palliative) solution to this problem is to replace significance testing with point estimates and

confidence intervals (e.g., Reichardt & Gollob, 1997) and effect sizes (e.g., Cohen, 1988). However, a strong counter-argument can be made that this strategy addresses only part of the problem (e.g., Krantz, 1999; Schmidt, 1996). A broader framework for overcoming these issues is to move beyond single-sample studies to the synthesis of findings drawn from multiple studies (e.g., Green & Hall, 1984; Hunter & Schmidt, 1996; Oakes, 1986; Schmidt, 1996).

Among several strategies, meta-analysis is a particularly important approach to research synthesis and great strides have been made over the past two decades in the development and application of these techniques within the social sciences (e.g., Cooper, Hedges, & Valentine, 2009; Schmidt & Hunter, 1977; Smith & Glass, 1977). Broadly defined, meta-analysis is a set of formalized procedures that allow for the synthesis of summary statistics drawn from a large number of existing studies. One of the original motivations for meta-analysis was that these techniques would further support the creation of a cumulative knowledge within the social sciences, particularly in psychology (e.g., Hunter & Schmidt, 1996; Schmidt, 1984). Indeed, Schmidt (1996) concludes "Unlike traditional methods based on significance tests, meta-analysis leads to correct conclusions and hence leads to cumulative knowledge" (p.119). There is no doubt that meta-analysis has substantially advanced our science toward this goal.

However, recall that one of the original motivations underlying the development of meta-analysis is the argument that single-study hypothesis testing has significantly impeded the progress of psychology. Yet the logical conclusion to this argument is that single-study analysis are useful only to the extent that they might represent a data point in some future meta-analysis. Further, the generalization of any single-study finding supported by traditional hypothesis tests may at best be irrelevant and at worst be impeding our progress as a science. So we find ourselves in the situation in which we have our dominant paradigm of hypothesis testing in single-sample analysis on one end, and the meta-analysis of a large number of studies on the other. And, at least in my own reading of the literature, we are each poked with a rather sharp stick to move to one of these two extremes.

But must we all be herded to one end of the pen or the other, or might there be options? Recent analytical and methodological developments reveal a progressively widening and highly captivating interior space that falls between the boundaries defined by single-study analysis and meta-analysis, a space upon which our field has yet to fully capitalize. More specifically, there are an increasing number of situations in which we may have access to the original raw data drawn from a modest number of individual studies that would be too small to support a meta-analysis but that might still be combined in some integrated and collaborative fashion to move meaningfully beyond single-study analysis. Could we in some way capitalize on this interior space to help us progress as a cumulative science while at the same time being cognizant of the very real practical issues that arise in applied research? The simple answer is yes. But how is such an integration best accomplished?

This seemingly innocuous question quickly morphs into a rather vexing problem. For example, could one simply fit models to data that have been pooled across two or more independent samples? How can this best be accomplished in a reliable and valid way? Would existing concerns about significance testing associated with single-sample analysis simply be generalized to the pooled data analysis? Alternatively, what if independent teams of researchers were pursuing similar programs of research at the same time? Could these be coordinated in a collaborative way to maximize the resulting scientific contributions? Further, might single-study analysis, pooled-study analysis and meta-analysis be conducted in way that would each support the other, thus allowing inferences that might not otherwise be possible? The goal of this special issue is to thoroughly explore these very questions.

This special issue consists of six manuscripts that each address a different dimension of methods for combining data in ways that move beyond single-study designs and help contribute to the pursuit of a cumulative psychological science. In the first paper, Andrea Hussong and I explore a variety of issues that arise when fitting models to data that has been pooled over two or more independent samples; we refer to this as integrative data analysis, or IDA. Next, Dan Bauer and Andrea Hussong propose a novel application of nonlinear factor analysis that allows for the estimation of measurement models for items that may be scaled differently and drawn from separate samples; they demonstrate this model using alcohol involvement measures pooled over two different studies. Jack McArdle, Kevin Grimm, Fumiaki Hamagami, Ryan Bowles and William Meredith then present a rigorous analysis of life span growth curves of cognition through the simultaneous estimation of item response theory models and latent curve models that are based on multiple scales drawn from multiple samples. Scott Hofer and Andrea Piccinin then describe the design and execution of a collaborative framework for building a cumulative science through the coordinated replication and integration of multiple studies and multiple data sets with a particular focus on research in aging. Next, Harris Cooper and Erika Patall provide a careful comparison of the relative advantages and disadvantages of current meta-analytic procedures with methods for combining raw data drawn from multiple studies and describe the conditions under which each approach may be beneficial and when they might be used in combination. Finally, Pat Shrout concludes with a thoughtful discussion of the core issues raised in the prior manuscripts and describes future directions for both quantitative and substantive researchers.

Taken together, this set of manuscripts explores the promising interior space that is defined by the single-study analysis on one extreme and the synthesis of summary statistics from a large set of studies on the other. Of course both single-study analysis and meta-analysis will always play a critical role as psychology continues to progress as a science. However, the following six papers reveal that much can be gained by navigating the area that lies between these two boundaries. Indeed, I believe that capitalizing on this middle-ground will help balance the lofty ideals of Don Quixote with the worldly practicality of Sancho Panza as we continue on our vitally important quest to build a truly cumulative psychological science.

Acknowledgments

This work was partially supported by grant DA013148 awarded to Patrick Curran and grant DA15398 jointly awarded to Andrea Hussong and Patrick Curran. I am indebted to Scott Maxwell for his unwavering support of this project; to Andrea Hussong and Dan Bauer for being argumentative; and to all of the contributing authors to this special issue for their unwavering dedication in the quest to improve our science of psychology.

References

- Bauer DJ, Hussong AM. Psychometric approaches for developing commensurate measures across independent studies: Traditional and new models. *Psychological Methods*. this issue.
- Cohen, J. Statistical power analysis for the behavioral sciences. 2nd. Hillsdale, NJ: Erlbaum; 1988.
- Cohen J. Things I have learned (thus far). *American Psychologist* 1990;45:1304–1312.
- Cohen J. The earth is round ($p < .05$). *American Psychologist* 1994;49:997–1003.
- Cooper, H.; Hedges, LV.; Valentine, JC. The handbook of research synthesis and meta-analysis. 2nd. New York: Russell Sage Foundation; 2009.
- Cooper H, Patall EA. The relative benefits of meta-analysis using individual participant data or aggregated data. *Psychological Methods*. this issue.
- Curran PJ, Hussong AM. Integrative data analysis: The simultaneous analysis of multiple data sets. *Psychological Methods*. this issue.
- D'Espagnat, B. Is science cumulative? A physicist viewpoint. In: Soler, L.; Sankey, H.; Hoyningen-Huene, PI, editors. *Rethinking scientific change and theory comparison: Stabilities, ruptures, Incommensurabilities*. Springer; 2008. p. 145–151.

- Green BF, Hall JA. Quantitative methods for literature reviews. *Annual Review of Psychology* 1984;35:37–53.
- Hofer SM, Piccinin AM. Integrative data analysis through coordination of measurement and analysis protocol across independent longitudinal studies. *Psychological Methods*. this issue.
- Howard, Don. Einstein, Kant, and the origins of logical empiricism. In: Salmon, W.; Wolters, G., editors. *Logic, Language, and the Structure of Scientific Theories*. University of Pittsburgh Press; Pittsburgh, PA: 1991. p. 45-106.
- Hunter JE, Schmidt FL. Cumulative research knowledge and social policy formulation: The critical role of meta-analysis. *Psychology, Public Policy, and Law* 1996;2:324–347.
- Krantz DH. The null hypothesis testing controversy in psychology. *Journal of the American Statistical Association* 1999;44:1372–1381.
- Kuhn, TS. *The structure of scientific revolutions: Third Edition*. Chicago: The University of Chicago Press; 1996.
- McArdle JJ, Grimm K, Hamagami F, Bowles R, Meredith W. Modeling life-span growth curves of cognition using longitudinal data with multiple samples and changing scales of measurement. *Psychological Methods*. this issue.
- Meehl PE. Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow process of soft psychology. *Journal of Consulting and Clinical Psychology* 1978;46:806–834.
- Mischel W. Our urban legends: Publishing. *APS Observer* 2008;9 21.
- Mischel W. The toothbrush problem. *APS Observer* 2009;11 21.
- Oakes, ML. *Statistical Inference: A commentary for the Social and Behavioral Sciences*. New York: Wiley; 1986.
- Reichardt, CS.; Gollob, HF. When confidence intervals should be used instead of statistical significant tests, and vice versa. In: Harlow, L.; Mulaik, S.; Steiger, J., editors. *What if there were no significance tests?*. Mahwah, NJ: Erlbaum; 1997. p. 259-286.
- Reisch GA. Did Kuhn kill logical empiricism? *Philosophy of Science* 1991;58:264–277.
- Schmidt, FL. Meta-Analysis: Implications for cumulative knowledge in the behavioral and social sciences; Invited address at the Annual Convention of the American Psychological Association; Toronto, Ontario, Canada. August 24-28; 1984.
- Schmidt FL. Statistical significance testing and cumulative knowledge in psychology: Implications for training of researchers. *Psychological Methods* 1996;1:115–129.
- Schmidt, FL.; Hunter, JE. Eight common but false objections to the discontinuation of significance testing in the analysis of research data. In: Harlow, L.; Mulaik, S.; Steiger, J., editors. *What if there were no significance tests?*. Mahwah, NJ: Erlbaum; 1997. p. 37-64.
- Shrout PE. Short and long views of integrative data analysis: Comments on contributions to the special issue. *Psychological Methods*. this issue.
- Smith ML, Glass GV. Meta-analysis of psychotherapy outcome studies. *American Psychologist* 1977;32:752–760. [PubMed: 921048]
- Weber, M. Commentary on “Is science cumulative? A Physicist viewpoint”, by Bernard D’Espagnat. In: Soler, L.; Sankey, H.; Hoyningen-Huene, Pl, editors. *Rethinking scientific change and theory comparison: Stabilities, ruptures, Incommensurabilities*. Springer; 2008. p. 153-156.