



eCOMMONS

Loyola University Chicago
Loyola eCommons

School of Education: Faculty Publications and
Other Works

Faculty Publications

2017

The Role of Theory in Quantitative Data Analysis

Terri D. Pigott

Loyola University Chicago, tpigott@luc.edu

Follow this and additional works at: https://ecommons.luc.edu/education_facpubs

 Part of the [Education Commons](#)

Recommended Citation

Pigott, Terri D.. The Role of Theory in Quantitative Data Analysis. The BERA/SAGE Handbook of Educational Research, , : 19, 2017. Retrieved from Loyola eCommons, School of Education: Faculty Publications and Other Works,

This Book Chapter is brought to you for free and open access by the Faculty Publications at Loyola eCommons. It has been accepted for inclusion in School of Education: Faculty Publications and Other Works by an authorized administrator of Loyola eCommons. For more information, please contact ecommons@luc.edu.

© Terri D. Pigott 2017

The Role of Theory in Quantitative Data Analysis

Terri D. Pigott

Theory and Data Analysis

If you take theory and models seriously, then (a) you need to elaborate clearly for yourself ‘what counts’ and how things supposedly fit together, and (b) you must hold yourself accountable to data.... From my perspective, theory is – or should be – the lifeblood of the empirical scientist. (Schoenfeld, 2010, p. 105)

The process of theorizing, collecting evidence, testing theory, revising theory, and then working through the cycle again is the basis of science. Theory is no less important when conducting a data analysis using quantitative methods. All statistical textbooks spend significant space on discussing the assumptions of statistical tests, the data requirements for a given estimation procedure, and the boundaries around the conclusions that can be drawn from results. Theory about the phenomenon that one is examining through quantitative data analysis is not only the driver for the methods used to collect evidence, but most importantly, the decisions made about how to model and test that theory.

This chapter discusses the importance of theory as the driver or ‘lifeblood’ of quantitative data analysis. All researchers, no matter how experienced, conduct research to contribute to

knowledge, and educational researchers, in particular, do so to improve the teaching and learning of all individuals in society. Using theory to guide and inform data analysis is one major strategy toward a research base that is reproducible and generalizable. Theory-driven analysis, as Schoenfeld (2010) reminds us, forces us to make our assumptions transparent, holds us accountable to the data, and also helps us to cumulate and develop knowledge to improve education.

While disciplinary research in fields such as psychology and sociology has traditionally been based in theory, Furlong (2013) outlines the pressures on current educational research to move the focus away from theory to more applied work that is directly applicable to practice. But as Lawn and Furlong (2009) describe, the disciplines and theoretically based research are important in the framing of educational research. Researchers in the qualitative tradition have more typically based their analyses on theoretical perspectives that they provide in a transparent way to readers. As Phillips and Burbules (2000) point out, the researcher using interpretive methods to describe human action needs to have a warrant, justified with evidence and theory, to back up their interpretation of an event. However, researchers using quantitative methods are less likely to provide clear theoretical perspectives that guide their analysis, spending more time discussing the rationale for particular modelling strategies than providing a clear warrant for their claims. This chapter makes a case for the importance of theory in quantitative research where researchers tend to highlight methods over theoretical frameworks.

What is Theory?

Theory, as defined by many dictionaries, is a verified description of relationships or underlying principles that describes a given phenomenon. A theory can help us to make sense of our observations, such as how the collective efficacy of an organization's members relates

to particular behaviours and attitudes of the organization's leader, or how increased testing standards for students might lead to higher academic performance. Theory is our working and somewhat tentative description of the principles that lead to the phenomenon we observe.

The type of theory that can guide and inform quantitative data analysis needs to be at a level where one can translate relationships between measured constructs into statistical models. A model in the sense I will use here is the embodiment of a theory (Schoenfeld, 2008), describing in testable terms the relationships among the measured constructs of the theory. For example, one can construct a simple model that hypothesizes a linear relationship or correlation between the measured efficacy of a school's teachers and the measured transformational leadership of the school's principal. The model is necessarily a simplification of a broader theory that uses imperfect measures of complex constructs like transformational leadership or academic performance to examine hypothesized relationships. It is likely that the relationships examined in any one study represent only small parts of a larger theory. Unless we have unlimited resources, we are unlikely to examine the entire path between a leader's transformational behaviours and an individual's teacher's sense of self-efficacy; instead we will examine whether the two measured constructs have a statistical relationship.

One example of the use of theory to guide data analysis is a study of how teacher–student interactions, students' social skills and classroom context predict children's emotional and behavioural difficulties in the classroom. Poulou (2014) draws on three theoretical approaches: a systems communication perspective on student–teacher interactions, social-emotional learning theory, and achievement goal theory. Poulou argues that these three theoretical perspectives complement each other, providing a more complete understanding of how emotional and behavioural problems may be exacerbated by a classroom's environment,

including how teachers and students interact. These three theories shape how the extensive survey and assessment data from a large set of Greek primary students are analyzed. She hypothesizes that a classroom goal structure that emphasizes mastery by providing students with authority over their learning, and meaningful tasks to complete, creates a more supportive environment for children. A classroom mastery goal structure and students' social skills determine the quality of the social relationships between teachers and students, and are predictive of the level of students' emotional and behavioural difficulties in the classroom. Poulou uses a structural equation model to test the inter-relatedness of these constructs as guided by theory.

A second example comes from Dumay and Galand's (2012) study of the impact of transformational leadership on teacher commitment. Dumay and Galand hypothesize that a teacher's organizational commitment is related to a number of factors, particularly a teacher's confidence that they can help students to learn (self-efficacy) and their perceptions of the climate of the school. Transformational leaders tend to nurture school cultures that build a sense of shared purpose among school community members and may be associated with teachers' self-efficacy. In their study, Dumay and Galand surveyed 660 teachers from 50 Belgian schools, collecting information about teacher's demographics and background as well as their self-efficacy, perceptions of school culture, and assessment of their principal's transformational leadership. Dumay and Galand did not test directly how a leader's transformative behaviours cause a teacher to have confidence in their ability to help students learn, only that the pattern of results is or is not consistent with the predictions from the theory. In fact, the data collected in the study were of teachers' perceptions of their leaders' behaviours rather than from direct observation of the leaders. Though theory guided the choice of the model examined, the model itself might be a simplification of the whole theory.

Though analytic models we build from theory cannot test all aspects of that theory, the theory we use for our models should guide the development of that model in terms of its statistical form and the predictions for the results. As Schoenfeld (2008) writes, three criteria for judging a model is its explanatory power, its predictive power and its scope. We can assess the adequacy of our theory by how well it helps us explain what we find in our analysis, how well it helps us predict what we found, and how well the theory generalizes to contexts outside of the original theory's scope. Our models should be inextricably tied to theory – not only in their development, but also in their estimation, and in their interpretation. Basing our models in theory then makes clear to our readers and to ourselves what parts of the model we are explicitly testing, and how that model relates to our initial theoretical assumptions. In addition, testing parts of the theory contributes to our development of better and more generalizable theory, seeing how far a theory can remain useful when applied in other contexts.

Importance of Theory to Guide Data Analysis

Theory should guide data analysis at all stages, beginning with the planning of the analysis, and ending at the interpretation of results. In the planning stages of a study, paying close attention to theory forces us to make our assumptions transparent. While conducting the analysis and examining preliminary results, theory can hold us accountable to data. Finally, using theory to help the interpretation of results highlights where our theory may work well, and places where the theory and/or model itself may need more careful study.

Planning the Data Analysis

Measures

Translating a theory into a statistical model requires the data analyst to make clear a number of assumptions. As Schoenfeld (2010) writes, the first stage is to elaborate what 'counts' and

how things fit together. In a statistical model, what 'counts' includes how we are measuring the important constructs from the model. In the Dumay and Galand (2012) study, a critical part of the translation from theory to model is the measurement of key constructs such as self-efficacy, transformational leadership and organizational commitment. The researcher needs to provide evidence that these measures of constructs in the theory are not only reliable, but valid measures of important constructs. Theory can help us to choose appropriate measures, and can also point out ways that our measures are not fully capturing the constructs. For example, Dumay and Galand raise a question about what is measured by teacher's ratings of principal's transformational leadership in terms of the model being tested. The leaders were not directly observed; instead, the study relied on the teachers' perceptions of their principal's behaviours. In this case, Dumay and Galand note that it is not clear if the theory does not adequately explain the results, or if the measures themselves are not adequate representations of the theory.

Making clear how the measures chosen fit with the theory also holds us accountable, as Schoenfeld (2010) suggests. There is currently much being written about research quality in medicine and the social sciences (for example, Ioannidis, 2005; Bakker and Wicherts, 2011; Simmons, Nelson and Simonsohn, 2011). In his examination of medical research, Ioannidis (2005) reaches the conclusion that many research findings are false due to a number of widespread, but questionable research practices. One of the issues he raises concerns the amount of flexibility in designs, definitions, outcomes and analytic modes in a given field. For example, for many drug treatments, the number of patients who survive after a given period is the accepted outcome, and can be measured with precision. However, in fields like education, there are many measures of constructs such as academic achievement or self-efficacy that could have varying levels of reliability and validity. In situations where there is

more flexibility in choosing measures, Ioannidis argues that researchers could choose less reliable measures that are biased and that lead to false results.

Making explicit how the choice of measures relates to the theory could guard against choosing measures that lead to the desired result. Simmons et al. (2011) demonstrate a number of practices that researchers use to find statistically significant results, including not reporting all measures collected, referred to in the literature as reporting bias. Pigott et al. (2013) have documented this practice in the educational research literature when comparing dissertations to their published versions. The most troubling case of outcome reporting bias was demonstrated by Vedula et al. (2009) where unpublished studies of the off-label indications from pharmaceutical companies were compared with their published versions. The published versions of these studies often changed the primacy of the outcome reported from the original study. This would be equivalent to an educational researcher examining reading achievement as a primary outcome in the original study, but then highlighting the results on attitudes toward reading instead when reporting on the study. Though being explicit about theory might not prevent researchers from biased reporting, it does provide a check on a researcher actively choosing measures with little connection to the theory guiding the study.

Choice of analytic models

Theory can also guide the researcher in choosing the form of the statistical models. Poulou uses a structural equation model to examine how student's self-reported social skills and the level of classroom autonomy and motivation relate to the level of emotional and behavioural difficulties in the classroom. For Dumay and Galand (2012), a key part of the theory includes attention to the multiple levels of schools where organizational factors such as overall school culture and the principal's leadership styles may be related to individual teachers' perceptions

of self-efficacy. Dumay and Galand use multi-level models to test their hypotheses about processes at different levels of the organization. In addition, Dumay and Galand also test whether the strength of the school culture mediated the relationship between perceptions of transformational leadership and teachers' organizational commitment. The framework of mediation and moderation analysis as discussed by Baron and Kenny (1986) provides an analytic model that can link directly to parts of theory. Finite mixture modeling (Gagné, 2006), and structural equation modeling (Hancock and Mueller, 2013) are other analytic frameworks that are also particularly well-suited for testing relationships between measured constructs that can be tied to theory. Just as a researcher uses a research question to guide the study design, theory should inform the data analyst when choosing the form of the statistical model to use in examining results.

Conducting the Data Analysis

As discussed above, many are concerned about research quality in medicine and the social sciences (for example, Ioannidis, 2005; Bakker and Wicherts, 2011; Simmons, Nelson and Simonsohn, 2011). In addition to flexibility around choice of measures and reporting bias, many of the practices called into question relate to suboptimal statistical practice, such as the search for statistical significance as the only goal of a study. Keeping the data analysis tied directly to a theory can serve as a barrier to engaging in bad statistical analysis.

Most students of statistics are told repeatedly that a significant p -value is not the ultimate goal of an analysis. However, pressures to publish and find a novel result (Ioannidis, 2005) may trump any admonition provided in a statistics class. There are a number of ways that the search for significant p -values can lead a researcher astray. One is the danger of conducting too many statistical tests, or the problem of multiplicity. The more statistical tests one conducts, the more likely one is to find a few significant p -values. One way to guard against

capitalizing on chance is to use methods for multiple hypothesis testing such as outlined by Keselman, Cribbie and Holland (1999). Another is to use theory to guide the number of statistical tests conducted. This approach is particularly important when a researcher is using large datasets to conduct research. For example, Poulou (2014) collected information from four different instruments from 962 students. A sample size this large will result in many statistically significant statistical tests, not all of which will be substantively important. Being explicit about how the statistical tests one conducts relates to part of the model and to the theory guiding the model can prevent a researcher from conducting too many tests. Planning a number of tests a priori, and explicitly reporting those tests allows the reader to make a judgement about whether the researcher capitalized on chance.

A related negative practice is what researchers such as Simonsohn et al. (2014) and Head et al. (2015) have called *p*-hacking, which occurs when researchers try out several statistical analyses and selectively report only those that are statistically significant. Head et al. examine the extent of this practice in the scientific literature using text-mining. They find widespread evidence of this practice, and variation among the disciplines in the extent of the practice. In response, Simonsohn et al. (2014) provides methodology to examine whether results in a single study are actually true, or whether they reflect selective reporting of the number of statistical tests conducted. As Head et al. (2015) suggest, one method for curbing this behaviour is to pre-specify the analyses in all reports that were planned a priori versus those that were conducted post hoc. In order to specify a priori hypotheses and analyses, a researcher needs to have a theory and associated model that can guide the choice of statistical analyses.

Another way to avoid selective reporting of only significant results is to stop the practice of null hypothesis testing altogether. Cumming (2014) calls for a ban on the use of statistical

tests in general, focusing instead on effect sizes, confidence intervals, and meta-analytic thinking, meaning placing one's current quantitative results in context with those from prior studies. Even if the field of educational research is not ready to ban significance testing, being transparent about how the statistical tests one is conducting relates directly to the model and the theory being tested can curb the over-reliance on finding statistical significance.

Interpreting the Results of Data Analysis

A priori theory

Once we have used theory to plan and keep us honest when conducting the analysis, it seems obvious that theory should guide the interpretation of the results. As researchers, we must remain attentive to theory at all stages and in particular when interpreting the results. One dangerous practice that a researcher might employ when interpreting results is what Kerr (1998) calls HARKing, or hypothesizing after results are known. As Kerr indicates, HARKing involves taking into account the post hoc plausibility of a hypothesis when deciding what hypotheses to advance when interpreting the results. In other words, HARKing occurs when a researcher is influenced by the results when deciding on the theory to use to explain those results. As Kerr points out, there are many incentives for researchers to engage in HARKing, such as the higher probability of successful publication when a study provides a clearer and more concise explication of results, that is, when results conform neatly to a theory. However, changing theory to fit data can lead an entire field down the wrong path. If the data does not fit one's a priori theory, the answer is not to change the theory but to report the finding that the theory does not predict the pattern of results obtained, and explore reasons why the theory failed.

Of course, using theory to guide the interpretation of a data analysis may be more complex and less straightforward than crafting a neat story about how the theory fits the data. Dumay

and Galand (2012) found mixed results for their hypotheses about how transformational leadership relates to teachers' organizational commitment. As they explain, the theory used in the study may not apply as well to loosely-coupled organizations such as schools; other applications of transformational leadership to organizational commitment were tested in more structured environments such as banks. Dumay and Galand do not reject the theory of transformational leadership guiding the study, but are exploring how well the theory might predict findings in contexts outside of those already tested.

Interpreting the novel result

Another problem identified in current critiques of research is the search for the novel or unexpected result. When interpreting results, we may be drawn to that unexpected finding, and as in Vedula et al.'s (2009) study, be tempted to re-frame the article to highlight that result. In active research areas, many researchers may be working simultaneously to find that one surprising or novel result. Ioannidis calls this the Proteus phenomenon, where one research team publishes a novel result and then another team finds the opposite. Ioannidis and Trikalinos (2005) demonstrate this phenomenon in the area of molecular genetics where research teams compete to find the most extreme results that are in turn more likely to be published. Using cumulative meta-analysis, a technique for tracking the effect size of studies across time, they found that in the field of molecular genetics, there is a rapid, early sequence of extreme results, followed by less extreme and more consistent results.

In educational research, the Proteus phenomenon might manifest in a search for a result that is unexpected or that runs counter to the prevailing theory and practice. Early evaluations of the preschool programme, Head Start in the US, found few positive results of the programme, contrary to what developers of the model expected. However, advances in evaluation methodology, statistical analysis and theory about effective early childhood

environments have tempered those early claims of no impact. Continued research has shown that early intervention can have significant influence on young children's development. A novel result itself must be considered novel or surprising because a theory and its associated model did not adequately predict the result. Theory can contextualize a novel result, and force the researcher to figure out where the assumptions of the model or theory fail to predict accurately, much as Dumay and Galand (2012) explain in their discussion.

Makel and Plucker (2014) come at the question of novelty from the perspective of replication. As they point out, there are few replications of experimental results in educational research, fewer even than in psychology. A consequence of the lack of replication is that we cannot know if a novel result will hold up in subsequent studies, and what the boundaries of that result are. The privileging of novel results over replicated ones could lead us to implementing ineffective or harmful treatments, or following paths of research that will eventually prove unworthy. Makel and Plucker distinguish between direct and conceptual forms of replication. While direct replication attempts to use the same methods as the original research, conceptual replication tests more general models based on the original research. Increasing the use of both direct and conceptual forms of replication would help a field distinguish true effects from novel but false results. Again, basing our analysis on theory, and thinking about conceptual replication can help us avoid the search for the novel result. We can instead ask ourselves how our results conform to our expectations based on theory, and explore potential reasons for the unexpected results.

Cognitive theory can also help us to use care in the interpretation of our data analyses. Kahneman (2011) talks about slow and fast thinking – about how we mostly use automatic, what he calls System 1 thinking, to make quick judgements about a situation. Kahneman argues that we are particularly bad at making statistical judgements. He outlines any number

of errors made when relying on our automatic System 1 such as seeing regularity produced by a random process, being unduly influenced by information most recently available, and regression to the mean. When conducting data analyses, we may also rely too heavily on quick thinking, and identifying patterns such as statistical significance, that may lead us astray. It takes complex thinking to examine a theory, break it down into testable parts, translate it into a testable statistical model, and then compare the results to what is predicted from the theory. It is much easier and more expedient to find the statistically significant result, and highlight that result in the reporting of the results. The slow-thinking approach – the asking of many questions, the going back and forth between the theory and the statistical model, the testing of alternative explanations for the findings – will ultimately lead to more defensible research results, and more progress in the field.

The Role of Theory-Guided Data Analysis in Producing Quality Research

When you are a quantitative researcher, it can be easy to get excited about estimating statistical models and finding statistical significance. Theory should, however, be foremost in our mind when developing our models, conducting the data analysis, and interpreting the results. The questions that guide educational researchers are important, and the quality of the research to improve the educational opportunities of all learners should be of highest priority. As discussed above, theory-driven data analysis forces us to make our assumptions clear to our audience, assumptions about how our choices about the development, testing and interpretation of our models relates directly to theory. Transparent and reproducible research is the hallmark of quality research, and helps us avoid the problems detailed by many in the literature (Ioannidis, 2005; Bakker and Wicherts, 2011; Simmons, Nelson and Simonsohn, 2011).

Theory-driven data analysis also has another benefit – contributing to the accumulation of knowledge. Theory builds a bridge between our work and those that came before us in the field. Drawing explicit connections to prior work through theory helps the field to see where our work fits. In a sense, using theory to connect our work to others' work is a form of conceptual replication (Makel and Plucker, 2014). The more we use theory to guide and inform our own data analyses, the more we can make explicit how our work tests our parts of a theory and begins to see where the theory holds and where it might need more evaluation.

Not all will agree with the emphasis on theory-driven data analysis. Bem (2000) argues that researchers should be free to explore their data to find interesting results, and that not all data analysis is confirmatory. Kuhn (2012) famously demonstrates that science may not progress linearly through experimentation and testing of theory but through disruptive breakthroughs that cause a field to radically change its ways of thinking. Despite these potential criticisms, using theory to guide key decisions in all stages of a data analysis does hold the researcher accountable for clearly articulating the connections between the theory, the model and the data. Poulou (2014) provides an example of how theory drives the analysis of large amount of survey and assessment data. Poulou collected data from 962 primary age children using four instruments measuring children's perceptions of their interactions with their teacher, self-reports on their social skills, their perceptions of their classroom's goals, and their strengths and difficulties with behaviours. This large amount of data is likely to yield many associations with statistically significant tests. However, Poulou uses a disciplined approach to examining her theoretical framework, rendering the choices made in the data analysis transparent to the reader. Others who then use her research can see what parts of the theory are tested in the model, and can build on her work to extend the knowledge in a field.

Conclusions

The quote that opens this chapter is from an article by Schoenfeld (2010) entitled, 'Reflections of an accidental theorist.' In this article, Schoenfeld provides examples of how theory frames his empirical work on the study of effective classroom teaching of mathematics. Though the examples he provides relate to qualitative research methods, his lessons are just as relevant to those of us engaged in quantitative data analysis. Schoenfeld (2010) writes that 'being explicit about theory and models helped me clarify what I was trying to understand and to test and refine my ideas' (Schoenfeld, 2010, p. 105).

Kuhn's (1991) later writing is consistent with Schoenfeld's iterative process around testing and then refining theories. In reconceptualising the historical philosophy of science, Kuhn (1991) asserts that what scientists produce and evaluate is not belief, but potential changes in belief. He writes that

what evaluation aims to select is not beliefs that correspond to a so-called real external world, but simply the better or best of the bodies of belief actually present to the evaluators at the time their judgements are reached. The criteria with respect to which evaluation is made are the standard philosopher's set: accuracy, breadth of application, consistency, simplicity.... (Kuhn, 1991, p. 119)

Thus, basing our data analysis on theory helps researchers to assess how well our theory explains some phenomenon and may eventually lead us change current beliefs and theories.

Theory-driven analysis contributes to the quality of research by making our assumptions and choices about the form of our models, the analysis of the data and the interpretation of our results clear and transparent. But, as Schoenfeld so eloquently demonstrates, continuing to cycle back and forth between our data analysis and our theory can also help us to understand our research more deeply. We are researchers because we want to contribute to

knowledge about how to improve education for all learners. Using theory to drive all stages of our analysis will enable us to produce quality research that others can use.

References

Bakker, M. and Wicherts, J. M. (2011). The (mis)reporting of statistical results in psychology journals. *Behaviour Research Methods*, 43(3), 666–678. doi:10.3758/s13428-011-0089-5.

Baron, R. M. and Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51(6), 1173–1182. doi:10.1037/0022-3514.51.6.1173.

Bem, D. J. (2000). Writing an empirical article. In R. J. Sternberg (Ed.) *Guide to Publishing in Psychology Journals* (pp. 3–16). Cambridge: Cambridge University Press.

Cumming, G. (2014). The new statistics why and how. *Psychological Science*, 25(1), 7–29. doi:10.1177/0956797613504966.

Dumay, X. and Galand, B. (2012). The multilevel impact of transformational leadership on teacher commitment: cognitive and motivational pathways. *British Educational Research Journal*, 38(5), 703–729. doi:10.1080/01411926.2011.577889.

Furlong, J. (2013). *Education: An anatomy of the discipline. Rescuing the University project*. London: Routledge.

Gagné, P. (2006). Mean and covariance structure mixture models. In G. R. Hancock and Mueller (Eds) *Structural Equation Modeling: A second course* (pp. 197-224). Greenwich, CT: Information Age Publishing.

Hancock, G. R. and Mueller, R. O. (2013). *Structural Equation Modeling: A second course*, 2nd edition. Charlotte, NC: Information Age Publishing.

Head, M. L., Holman, L., Lanfear, R., Kahn, A. T. and Jennions, M. D. (2015). The extent and consequences of p-hacking in science. *PLoS Biol*, 13(3), e1002106.

doi:10.1371/journal.pbio.1002106.

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Med*, 2(8), e124. doi:10.1371/journal.pmed.0020124.

Ioannidis, J. P. A. and Trikalinos, T. A. (2005). Early extreme contradictory estimates may appear in published research: the Proteus phenomenon in molecular genetics research and randomized trials. *Journal of Clinical Epidemiology*, 58, 543–549.

doi:10.1016/j.jclinepi.2004.10.019.

Kahneman, D. (2011). *Thinking, Fast and Slow*, 1st edition. New York: Farrar, Straus and Giroux.

Kerr, N. L. (1998). HARKing: hypothesizing after the results are known. *Personality & Social Psychology Review (Lawrence Erlbaum Associates)*, 2(3), 196.

Keselman, H. J., Cribbie, R. and Holland, B. (1999). The pairwise multiple comparison multiplicity problem: an alternative approach to familywise and comparisonwise type i error control. *Psychological Methods*, 4(1), 58–69.

Kuhn, T. S. (1991). The trouble with the historical philosophy of science. In J. Conant and J. Haugeland (Eds) (2000), *The Road Since Structure* (pp. 105–120). Chicago, IL: University of Chicago Press.

Kuhn, T. S. (2012). *The Structure of Scientific Revolutions: 50th anniversary edition*. Chicago, IL: University of Chicago Press.

Lawn, M. and Furlong, J. (2009). The disciplines of education in the UK: between the ghost and the shadow. *Oxford Review of Education*, 35(5), 541–552.

Makel, M. C. and Plucker, J. A. (2014). Facts are more important than novelty replication in the education sciences. *Educational Researcher*, 43(6), 304–316.

doi:10.3102/0013189X14545513.

Phillips, D. C. and Burbules, N. C. (2000). *Postpositivism and Educational Research*.

Lanham, MD: Rowman & Littlefield.

Pigott, T. D., Valentine, J. C., Polanin, J. R., Williams, R. T. and Canada, D. D. (2013).

Outcome-reporting bias in education research. *Educational Researcher*, 42(8), 424–432.

doi:10.3102/0013189X13507104.

Poulou, M. (2014). The effects on students' emotional and behavioural difficulties of teacher–student interactions, students' social skills and classroom context. *British Educational Research Journal*, 40(6), 986–1004. doi:10.1002/berj.3131.

Schoenfeld, A. H. (2008). Chapter 2: On modeling teachers' in-the-moment decision making. *Journal for Research in Mathematics Education. Monograph*, 14, 45–96. Retrieved on 9/15/2015 from www.jstor.org/stable/30037741.

Schoenfeld, A. H. (2010). Reflections of an accidental theorist. *Journal for Research in Mathematics Education*, 41(2), 104–116.

Simmons, J. P., Nelson, L. D. and Simonsohn, U. (2011). False-positive psychology undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22(11), 1359–1366. doi:10.1177/0956797611417632.

Simonsohn, U., Nelson, L. D. and Simmons, J. P. (2014). P-curve: a key to the file-drawer. *Journal of Experimental Psychology: General*, 143(2), 534–547. doi:10.1037/a0033242.

Vedula, S. S., Bero, L., Scherer, R. W. and Dickersin, K. (2009). Outcome reporting in industry-sponsored trials of gabapentin for off-label use. *New England Journal of Medicine*, 361(20), 1963–1971. doi:10.1056/NEJMsa0906126.