



Psychological Inquiry

An International Journal for the Advancement of Psychological Theory

ISSN: (Print) (Online) Journal homepage: <https://www.tandfonline.com/loi/hpli20>

The Struggle Is Real: Challenges and Solutions in Theory Building

Andrew R. A. Conway , Kristof Kovacs , Han Hao , Sara A. Goring & Christopher Schmank

To cite this article: Andrew R. A. Conway , Kristof Kovacs , Han Hao , Sara A. Goring & Christopher Schmank (2020) The Struggle Is Real: Challenges and Solutions in Theory Building, Psychological Inquiry, 31:4, 302-309, DOI: [10.1080/1047840X.2020.1853468](https://doi.org/10.1080/1047840X.2020.1853468)

To link to this article: <https://doi.org/10.1080/1047840X.2020.1853468>



Published online: 07 Jan 2021.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)



The Struggle Is Real: Challenges and Solutions in Theory Building

Andrew R. A. Conway^a, Kristof Kovacs^b, Han Hao^a, Sara A. Goring^a, and Christopher Schmank^a

^aClaremont Graduate University, Claremont, California, USA; ^bEotvos Lorand University, Budapest, Hungary

Strong theories are sorely lacking in the applied social sciences, especially in psychology. Elko Fried identifies fundamental problems that are common in social science research and explains how these problems manifest themselves in the literature, impede scientific progress, and contribute to the lack of theory building. He covers a lot of ground: weak theories, the conflation of theoretical and statistical models, problematic inferences, equating latent variables with psychological constructs, and more. As intelligence researchers, we are all too familiar with these kinds of problems; they have been sources of confusion and barriers to progress for over a century. A decade ago, the continued lack of concern over these issues in the field of intelligence became our major concern, and our motivation to develop a new approach. Suffice to say, Elko Fried is music to our ears.

It was indeed a pleasure to read the target article and we appreciate the opportunity to respond in a commentary. We hope to share some insights from our program of research on working memory and intelligence. In our work, we address several of the problems discussed by Fried, and we proposed a solution, Process Overlap Theory (POT), a new approach to intelligence that integrates evidence from cognitive psychology, psychometrics, and neuroscience (Kovacs & Conway, 2016, 2019a). POT is largely motivated by research on working memory and so our commentary will draw on examples from both the intelligence literature and the working memory literature.

The commentary is divided into three main sections. In the first section, we review several points of agreement with Fried, drawing on examples from our research. In the second section, we discuss a few of our concerns. For the record, we agree with Fried much more than we disagree. We devote more attention to the points of disagreement here in the commentary because we see this as a unique opportunity to provide constructive criticism and promote discussion. In the third section, we discuss problems with graduate training. We argue that most graduate programs lack the kind of formal training that is necessary to promote theory building. The lead author (Conway) teaches graduate-level statistics and three of the coauthors (Hao, Goring, Schmank) are current graduate student so we offer some advice on how to address this problem.

Common Ground

The Fried article and our recent work on intelligence share a lot of common ground, which is pretty remarkable given that we come from different backgrounds and work in different disciplines. For the record, we are all cognitive psychologists, trained in experimental psychology, but we also study individual differences in cognitive abilities and intelligence. This means that our initial training in graduate school had a particular focus, and then we pursued supplementary training. For example, in graduate school, the focus of research methods was experimental design, the main approach to statistics was the general linear model and NHST, and “models” were either cognitive or computational. We all supplemented that training by learning advanced statistics and psychometrics. For example, three of us are current graduate students (Hao, Goring, Schmank) and elected to take several advanced statistics courses, including factor analysis, structural equation modeling, multilevel modeling, and item response theory.

Here we touch on four sections of the Fried paper where we discovered common ground: (1) Need for theory; (2) The conflation of theoretical and statistical models; (3) Problematic inferences from the factor literature; and (4) Compatibility between theories and models.

Need for Theory

Maybe we are reaching the day of the theorist in psychology, much as it exists in other sciences such as physics.—
Newell (1973)

If you ask a cognitive psychologist about the importance of theory, chances are they will cite Alan Newell’s classic, “You Can’t Play 20 Questions with Nature and Win.” Newell was invited to comment on a series of papers presented at a symposium and in his attempt to “put them all together” he became “distressed” by their lack of coherence:

But as I tried to put them all together, I was led back from the particular results described to a set of results that these papers referenced and used, in a qualitative sort of way. These led me back to yet other papers, many by the same group of authors and of equal merit and precision. It became less and less clear to me that all these papers were cumulating. Only the barest fraction of each prior paper found its way into the next (though fortunately there were some exceptions), and these experiments I was considering (those today) seemed destined to play a similar role vis a vis the future.

Newell presents an entertaining and compelling argument for the value of theory. In the ideal scenario, a strong theory makes predictions, which guides research, and in turn generates results and papers that become cumulative, and science advances. No doubt, there is a need for more theory in the social sciences and we should all value strong theories.

Another benefit of strong theories is the potential for theory integration. Our work on POT is a good example. POT is a theory about intelligence that is largely motivated by research on working memory. Developing the model was basically an exercise in theory integration. POT integrates sampling theory from the field of psychometrics (Thomson, 1916), the “systems approach” to the study of intelligence (Detterman, 1987), and working memory theories from cognitive psychology (Cowan, 2005; Engle, 2002; Oberauer, 2009). According to sampling theory, the general factor of intelligence (g) does not represent a common cause. Instead, g emerges from a pattern of positive correlations across tests. The correlations across tests are caused by an overlap of cognitive processes sampled by the tests. The fundamental premise of POT is that individual items on tests of intelligence are likely to require multiple cognitive processes, some domain-general and some domain-specific. Crucially, there is no one single process that is required by all items on every test. In other words, there is no general ability. The systems approach is similar; it argues that intelligence is composed of multiple interrelated faculties, some of which are more central than others.

Next, to provide an account of the domain-general and domain-specific processes, POT incorporates theories of working memory. Working memory is a system that maintains information in a readily accessible state while processing new information and avoiding interference. According to both Cowan (2005) and Oberauer (2009), working memory involves multiple cognitive processes, some associated with short-term memory and some associated with attention. In POT, we don’t introduce a new account of what these processes are, we simply incorporate ideas from Cowan and Oberauer. Also, Engle’s theory (2002) claims that individual differences in working memory capacity are largely due to executive attention processes, and POT incorporates that view as well.

And then, of course, the theories of working memory that informed POT were informed by prior theories. As mentioned, working memory involves both memory and attention. To give one example, Oberauer’s theory of working memory integrates the work of Halford and colleagues (Halford, Wilson, & Phillips, 1998) to provide an account of memory mechanisms and the work of Norman and Shallice (1980) to provide an account of attentional processes.

So, we agree that theory is important, but placing too much value on theory can be a problem. In our experience, this is common in Psychology, especially in the United States. At some top tier programs, job candidates for an Assistant Professor position are expected to already have their own theory of X, Y, or Z. We refer to this as theory worship and it results in more weak theories, as well as niche theories.

Theory worship is a problem because placing too much value on theory means that too little value is placed on empiricism. As a result, expertise in measurement and sophisticated approaches to data collection are not considered to be as important or impressive as having a niche theory. In Psychology, there is evidence of this problem in the type of papers that are published in top tier journals, the grants that are funded, and the people who get tenure track jobs.

Again, we agree that it is important to value theory, but we also admire and respect our fellow scientists who adhere to old fashioned “dust bowl empiricism” and maintain that the collection of data should be valued above any particular content. We see merit in both approaches and we strongly believe that it is important to reward and incentivize both and equally. Psychology rewards and incentivizes scientists for theory building more than it does for empiricism, which helps to explain why it is a science cluttered with weak theories, niche theories, and bad measurement.

The Conflation of Theoretical and Statistical Models

When cognitive psychologists read the intelligence literature, this is probably the biggest source of confusion. We agree with Fried that the word “model” is part of the problem. In cognitive psychology, a “model” typically refers to a cognitive or computational model. These models provide an account of the cognitive processes underlying some aspect of cognition, or task performance. Most cognitive models are normative models, that is, they provide an account of typical behavior but they do not provide an account of individual differences. In cognitive psychology, there is generally good correspondence between theory and model but the scope of theory is often limited to a normative account of a very specific aspect of cognition.

In psychometrics, “model” typically refers to a statistical model. Here, conflation between theories and statistical models is rampant. Many theories of intelligence are really just statistical models; most commonly, a latent variable model with a hierarchical reflective general factor or a bifactor model with a general factor. These models provide an account of the covariance structure of test scores but they do not provide an account of the cognitive processes required to perform the test.

To add to the confusion, there are also cognitive theories that are really just statistical models. For example, Engle’s (2002) executive attention theory of working memory is not the same type of theory as the ones proposed by Cowan (2005) or Oberauer (2009), or the original theory of working memory developed by Baddeley and Hitch (1974). Engle’s theory provides an account of individual differences in working memory capacity, but technically speaking, it does not provide an account of the cognitive processes involved in task performance. More specifically, the theories proposed by Cowan, Oberauer, and Baddeley and Hitch make different predictions about specific aspects of the cognitive processes associated with working memory, whereas Engle’s theory makes predictions about the covariance structure of

various measures of individual differences in cognitive ability.

Problematic Inferences from the Factor Literature

Of course, we strongly agree with Fried that problematic inferences from the factor literature are a source of confusion. This problem originated in the field of intelligence with the work of Spearman (1904). Spearman notoriously interpreted the general factor to be a common source of test variance, reflecting a psychological attribute. For over 100 years, this has been the dominant interpretation of the general factor in Psychology. And still, there is no coherent theoretical account that incorporates an interpretation of the general factor as a general cognitive ability. Talk about a problematic inference from the factor literature!

Fried explains why this type of inference is problematic: “the statistical model does not provide conclusive information about the data generating (i.e., causal) mechanism” (p. 275). Exactly. To illustrate this problem with respect to Spearman’s *g*, we recently conducted a simulation where test scores were generated for 9 different tests considered to be measures of 3 broad cognitive abilities (spatial ability, verbal ability, and fluid reasoning). The test scores were generated according to a mathematical formulation of POT, which is expressed as an item response model. In brief, the probability of a person responding correctly to an item on a test is determined by the person’s “ability” on multiple cognitive processes, some of which are domain-general and some domain-specific. Each individual test item samples a different set of cognitive processes. We assumed 200 different processes (50 verbal, 50 spatial, 50 fluid, and 50 domain-general executive attention processes). We assumed a large sample of people ($N=1,000$), which means that 200 “abilities” (one for each process) were randomly generated for each person (all abilities were normally distributed). POT is based on a sampling algorithm where the probability of sampling a domain-general process is *greater than* the probability of sampling a domain-specific process on fluid reasoning tests, but for spatial and verbal tests the probability of sampling a domain-general process is *less than* the probability of sampling a domain-specific process. Test scores generated from the POT algorithm were submitted to a confirmatory factor analysis and a hierarchical three-factor model provided an excellent fit to the data. This is proof of principle that a latent variable model with reflective *g* can be observed in the absence of a general ability parameter (Conway, Kovacs, Hao, & Snijder, 2019; Kovacs & Conway, 2019b).

Thus, POT rejects the idea that the general factor is a common cause or a psychological trait. This change in the interpretation of the general factor has major implications. First, we avoid Spearman’s problematic inference. Second, if the general factor does not represent a psychological trait then there is no such thing as “general intelligence” or “general cognitive ability.” This means that research programs aimed at identifying the cognitive correlate of *g*, or the neural substrate of *g*, or the genetic contribution to *g* are

all off the mark. Instead of a focus on general intelligence, research should be aimed at understanding broad cognitive abilities, such as spatial ability, verbal ability, and fluid reasoning, as well as underlying cognitive processes, such as working memory, executive attention, and processing speed (Kovacs & Conway, 2019b).

Unfortunately, Spearman’s problematic inference is being repeated in cognitive psychology research on individual differences in working memory. Multiple studies now show that when a large battery of working memory tasks is administered to a large group of subjects, the positive manifold is observed, just like intelligence. And just like batteries of intelligence tests, patterns of convergence and divergence appear among the correlations. Working memory tasks with verbal content tend to be more strongly correlated with other verbal tasks than with spatial tasks. This has led to a debate in the field about the domain-generality of working memory capacity (Kane et al., 2004). In our view, this is history repeating. In the intelligence literature, Spearman and Thurstone had this same debate; Spearman argued for the importance of general ability whereas Thurstone downplayed the importance of general ability and argued for the importance of several distinct “primary abilities.” In the end, the two sides acknowledged that both general ability and a set of primary abilities are important. Likewise, in the cognitive psychology literature, it is clear that both domain-general and domain-specific aspects of working memory are important.

In fact, POT offers a unique perspective on this issue. According to POT, domain-general processes associated with executive attention serve as a bottleneck on performance and poor executive attention constrains cognitive ability in a domain-general fashion. POT, therefore, predicts that cross-domain correlations (verbal, spatial) will be higher for individuals with poor executive attention and lower for individuals with greater executive attention. This pattern of correlations is consistent with the idea of ability differentiation in intelligence. Indeed, there is evidence for ability differentiation in tests of intelligence and in our recent work we show evidence of ability differentiation in working memory (Kovacs, Molenaar, & Conway, 2019).

Compatibility Between Theories and Models

Taken together, POT resolves many of the problems discussed by Fried. POT provides a novel account of empirical evidence, which is explained verbally, and formally in a mathematical model based on item-response theory. POT makes a clear distinction between the corresponding cognitive model and statistical model (which is a psychometric model). Moreover, POT reinforces the value of strong theories by incorporating prior theoretical approaches, building a cumulative science, and contributing to scientific progress. Until very recently, however, there was still a fundamental problem with POT; the theory and the psychometric model were incompatible. According to POT, the general factor of intelligence is not a common cause, does not reflect a psychological trait, and should not be interpreted as a general

cognitive ability. A latent variable model with a reflective general factor is therefore incompatible with POT. Initially, we didn't address this problem of incompatibility. Thanks to Fried and others and their recent work on psychometric network modeling, we now have a better option. In our recent work, we compared latent variable models and network models of data from the WAIS (a standardized battery of intelligence tests) and found that network models provided a better fit to the data (Schmank, Goring, Kovacs, & Conway, 2019). To be clear, our preference for the network model is not really about model fit; we prefer the network model because it is compatible with POT and reflective latent variable models are incompatible.

Still, our network model of the WAIS is not ideal. According to POT, cognitive tasks are not processed pure. This means that the correspondence between nodes in the model and cognitive processes in the theory is not complete. An exciting new development is an approach called "cognitive psychometrics" (or computational psychometrics). In this approach, raw scores on a task are not used as measures of individual differences. Instead, the indicator variables are parameters generated from a computation model of task performance. This assumes, of course, the availability of established computational models of cognitive tasks. That will require more work. But eventually, if this approach is successful, then the nodes in a network model will more directly correspond to the cognitive processes specified in POT.

Concerns

In this section we discuss three main concerns with the Fried target article; (1) lack of clarity with respect to weak vs. strong theories; (2) framing of factor modeling and network modeling as two disciplines; (3) causality issues with network models.

Weak Vs. Strong Theories

As a key point in his argument, Fried differentiates strong and weak theories: "I define weak theories as narrative and imprecise accounts of hypotheses, vulnerable to hidden assumptions and other unknowns. They do not spell out the functional form in which two variables relate to each other, the conditions under which a hypothesized effect should occur, or the magnitude of a proposed effect. It, therefore, remains somewhat unclear what the theory actually explains or predicts, or how to use the theory for purposes of control (such as informing treatments in clinical psychology)" (p. 272). It is also claimed that weak theories "provide neither precise explanations nor predictions, and it can be difficult to find out if data support weak theories or not."

Given these descriptions, we would find it difficult to teach a neural network to tell weak and strong theories apart. It seems clear that a chief characteristic of a weak theory is a lack of mathematical formalization ("the functional form in which two variables relate to each other"). Yet the two quotes above differ in important ways. The former one allows weak theories to provide predictions, albeit "unclear"

ones, i.e., leaving the conditions under which effects occur or the magnitude of such effects unspecified. The second quote implies that weak theories do not provide predictions that would allow data to support them.

For instance, despite all of our common ground, we were unable to decode whether POT is a weak or a strong theory. In Fried's paper, POT is referenced after the claim "numerous theories persist that are underdetermined by data," (p. 281) and POT does not appear on the list of examples of strong theories. We would like to believe that POT is a strong theory, as it is formalized as an IRT model and provides predictions, some of which have already been supported (Kovacs, Molenaar, & Conway, 2019). But it does not provide predictions with regard to the magnitude of effects, and its formalization relates to many more than two variables.

More generally, even though formalization indeed makes psychological theories more alike to the ones in the harder sciences, it does not seem necessary for a psychological theory to be formalized—or even be formalizable—to provide falsifiable predictions, i.e., to be scientific. Take, for example, a particular instance of theories of social learning in very young children. An influential theory claims that children learn through imitating adults, without inferring whether a given action is necessary toward achieving a given goal: if adults do it, children automatically imitate it. In an experiment involving 14-month-olds a demonstrator illuminated a light-box by leaning forward and touching it with her forehead. A week later two-thirds of the children participating in the experiment reenacted this action: they illuminated the light-box with their forehead, even though they could have used their hands, which appears to be a rather convincing piece of evidence in support of the theory (Meltzoff, 1988).

A competing theory claims that children even this young are capable of taking an "intentional stance," i.e., to attribute intentions to agents and presume that their behavior is goal-directed when interpreting their behavior (Gergely, Nádasdy, Csibra, & Bíró, 1995). Proponents of this theory interpreted Meltzoff's results with a twist and proposed that children did not, in fact, *automatically* imitate the so-called "head action." Rather, since they interpret the behavior of others as rational and goal-directed, they probably inferred that since the demonstrator obviously could have used her hands, there *must be a reason* why she used her head instead. That is why they imitate the behavior exactly as they have seen it.

So far these two theories are underdetermined by the results of Meltzoff's original experiment. Therefore the proponents of the second approach replicated the experiment under two conditions. Under the first condition, they repeated the original experiment exactly as it took place. But in the second condition, the demonstrator pretended to be cold and wrapped a blanket around herself, which she held with both hands, making her hand occupied when she touched the light box with her forehead.

Under this condition, the two theories provide different predictions. If children are automatic imitators then no

difference is expected between the two conditions. But if children attribute rationality to agents, then they will repeat the head action a week later in the original condition only. Since in the “hands constrained” condition it is clear why the demonstrator used her forehead, children will use their hands instead. Results supported the second theory; in the hands-free condition, 69% of children reenacted the head action, almost exactly as in the original study. But under the hands occupied condition only 21% did so, 79% used their hands.

Both theories are weak theories according to Fried’s categorization; they do not provide a mathematical formalization, nor do they predict the magnitude of effects. Yet they describe a matter of great importance (whether young children attribute rationality to adults when they learn from them), and an elegant experiment could be devised based on the different predictions of the two theories. Moreover, we see no obvious way how the proposals of the theories (automatic imitation vs. attribution of rationality) could be mathematically formalized, let alone predict the magnitude of effects. Hence, even though we agree that formalization is useful whenever possible and suitable to the given area of inquiry, we are afraid that the categorization Fried proposes would throw much of the scientific baby out with the unscientific bath water. Overall, the very distinction between weak and strong theories offered by Fried comes across as a weak one. In our view, the focus should be on providing testable predictions rather than formalization or precisely predicting the magnitude of effects.

Factor Models Vs. Network Models

Fried provides a thorough and comprehensive summary of how the essential assumptions of factor models and network models lead to corresponding statistical inferences, as well as how such statistical inferences can be conceptualized or mis-conceptualized into not only statistical but also theoretical claims. Specifically, for a common reflective factor model, it is assumed that the latent factor being extracted from the communality of all manifest variables is the identified common cause for all the observed correlations among manifest variables (Kievit et al., 2011). On the other hand, network models focus on direct inferences to conditional dependence among variables, instead of imposing indirectly estimated latent factors (Fried, 2020). While it is of great importance to acknowledge and identify the divergence of latent factor modeling and network modeling and their corresponding theoretical-statistical inference gaps, the two branches of statistical modeling approaches are not necessarily mutually exclusive: they are different statistical approaches to the data, but are no different in filling the statistical-to-theoretical gap. After all, both of them provide only descriptions to correlational data and eventually describe theories, and neither of them could be used as the only evidence to “validate” a theory. Thus, it does not mean that one can always surpass the other at every aspect of research practice.

Here we illustrate the argument using the role of the general factor of intelligence (g) in POT as an example. Again, POT interprets g as a result of sampling, where multiple cognitive processes are involved in the performance of each test, and the set of processes required to complete a battery of tests is sampled in an overlapping manner across tests. Importantly, none of the cognitive processes is exclusively involved across the entire battery of tests. Therefore, from a theoretical perspective, a network model is more compatible with POT than a latent variable model, such that the g factor is not assumed to be the cause of the covariance among intelligence tests but is assumed to be caused by an overlap of processes tapped by tests.

However, this does not indicate that a network model can be used to validate POT by itself. For a set of behavioral observations of intelligence (test scores), a network model, with nodes as subjects’ performance on specific tests and edges as their relationships, is more compatible with POT than a reflective latent factor model with manifest variables being test performances and factor loadings being their reliabilities of reflecting the latent construct (intelligence). Still, the estimated network model does not represent the underlying cognitive mechanisms that POT describes, because the behavioral observations are also not direct representations of the sampled cognitive processes, and the between-test conditional relationships revealed by network models do not reveal the between-process conditional relationships. For example, in our recent simulation study, which we described earlier, we demonstrated that g can emerge from the covariance among test scores in the absence of a general cognitive ability. The simulation also demonstrated that g can emerge in the absence of conditional dependence of the sampled processes; the tests scores were generated from 200 *orthogonal* ability measures. In this case, the corresponding network model estimated from the simulated test scores cannot capture the orthogonality of the simulated cognitive processes, and cannot capture the sampling mechanism, although a network model on the simulated data aligns with the theoretical model of POT better than a latent factor model.

The compatibility of network modeling and POT also does not suggest that performance on intelligence tests should *never* be analyzed and described with latent factor models. It is well established that g is a strong predictor of job performance and academic achievement (e.g., Kuncel, Hezlett, & Ones, 2004). Therefore, a formative latent factor model of g is still a useful tool when using a specified set of indicators (intelligence tests) to predict outcomes of interest. In a formative latent factor model, g is interpreted as an index of the sampled indicators of intelligence, which is dependent on populations and tasks, instead of being assumed as a psychological construct (van der Maas, Kan, & Borsboom, 2014).

Causality

Fried compares network models and factor models in terms of the possibilities of causal analysis and claims that network analysis provides an alternative to factor models by replacing

a latent common cause; in a network causality manifests itself through the connections between nodes. In other words, latent variable modeling and network modeling offer latent versus manifest causal structures, respectively: “Network psychometric models estimate conditional dependence relations among variables with the goal to guide causal inference,” and “Network theory suggests that correlations (...) stem from causal interactions between items rather than from one shared origin [as in factor models].”

We perceive the issue of causality to be more complex than that. Technically, exploratory factor analysis (EFA) is simply a data reduction technique—as acknowledged by Fried. Confirmatory factor models do indeed allow for the testing of causal structures, however, they do not *establish* causal connections from covariance structures. “SEM [structural equation modeling] is an inference engine that takes in two inputs, qualitative causal assumptions and empirical data, and derives two logical consequences of these inputs: quantitative causal conclusions and statistical measures of fit for the testable implications of the assumptions. Failure to fit the data casts doubt on the strong causal assumptions of zero coefficients or zero covariances and guides the researcher to diagnose or repair the structural misspecifications. Fitting the data does not “prove” the causal assumptions, but it makes them tentatively more plausible” (Bollen & Pearl, 2013, p. 309).

Similarly to EFA, a network model is a representation of a correlation matrix that simplifies it and enables an easier access to the patterns of covariance structure, but unlike EFA, it does so by enabling visual access to such patterns rather than reducing data. (Unless there is a threshold below which correlations are not visualized but this does not provide additional information, in fact it is the equivalent of deleting all values below the same threshold from the correlation matrix in order to make it more accessible.) “Indeed, the network model is basically an alternative representation of the correlation matrix and the saturated model accounts for (i.e., is identical to) all of the covariance in the matrix” (McFarland, 2020, p. 2).

Overall, while network theory does differ drastically from latent variable modeling in the sense that possible causal relations are embedded between elements rather than the result of a common cause, network models do not *establish* causality any more than factor models do.

Graduate Training

Regardless of subfield, graduate students in psychology should leave their programs with knowledge of the concepts relevant to their research area and extensive experience designing and conducting research. We have isolated four areas all graduating students should have mastery of (1) proper research techniques and methodology; (2) implementing rigorous data-diagnostics and conducting appropriate and robust statistical analyses; (3) an understanding of the tenets of the philosophy of science and how to apply those principles in their own research; and (4) experience properly developing and evaluating theoretical frameworks.

Considering the first two points, most graduate programs in psychology require the completion of courses in methodology and statistics, but there are aspects that could be improved upon to be more effective. Yet, some graduate schools in psychology neglect the latter two points, despite being imperative for the advancement of psychological research.

First, a rigorous and comprehensive understanding of the research techniques and experimental methodologies used by researchers in one’s subfield is required. It is imperative that the methodological courses instill a working knowledge of the importance of appropriate psychometrics to ensure that experimental research conducted by future scientists will be methodologically sound and generalizable. However, topics like validity and reliability tend to be broadly covered in psychology courses from a conceptual perspective, but not enough attention is given to teaching students how to apply these concepts in a technical or applied manner. Many students do not fully grasp how to interrogate if an operational definition of a variable is accurately characterized by the psychometric instrument selected to evaluate the construct. For example, little focus is given to explicitly defining the boundaries of an operational definition for a construct, such that it is delineated from similar concepts and their associated measures. Furthermore, not enough technical training is offered to graduate students for formally evaluating whether a measure is truly valid and reliable. Although many students can define the relevant terms on a test (e.g., *what is construct validity?*), it is commonplace for graduate students to not know how to apply these concepts to the measures used in their research.

Next, connecting data collected via methodology with robust statistical analyses that are aligned with specified research questions is necessary. Unfortunately, many graduate researchers simply conduct analyses they are familiar with and are never shown how to weigh the limitations and assumptions of different analyses, in order to select the best option given the research question and data. Relatedly, many students are not shown the importance of rigorous data-cleaning and assumption-checks, or that decisions during cleaning and analyzing of data should be explained and justified explicitly. The subjective nature of data-cleaning and designing analytical methodology requires researchers to be transparent and able to defend the choices made when collecting and analyzing the data. However, applied graduate programs have often neglected to facilitate graduate students exploring analytical and methodological concepts for their own intellectual curiosity and research development. In parallel fashion, many applied psychology graduate students do not embrace these concepts as foundational learning and prefer to be told hard-and-fast rules for making all methodological and analytical decisions. As a result of this disconnect, there are many published examples of inappropriate data/non-replicable models, incongruous pairings between research question and analysis, and a dearth of creative analytical approaches in the field. Although the issues concerning research methodology and statistics must be addressed, many of the knowledge gaps among graduate

students are symptoms of a much deeper problem: misinterpretation of the foundations of scientific inquiry.

The shortcomings of many psychology graduate programs extend beyond the consideration of statistics and methodology. In fact, many researchers have called for a stronger effort and commitment from institutions to (a) effectively impart the foundational tenets of the philosophy of science to students in an accessible and thorough manner; (b) offer formal instruction for constructing theoretical frameworks, evaluating competing theoretical perspectives, and how to infer knowledge from theory. Simply providing students with a general understanding of methodology and statistics is not enough; these factors must be grounded in an appreciative understanding for how scientific discovery is brought about, how it is organized, and how it is disseminated within the field. It is imperative to go beyond merely paying lip-service to the tenets underlying the scientific method and teach these concepts in a more deliberate way. Specifically, the four tenets of science that must be given greater focus in psychology are: (a) replicability, (b) transparency and public verifiability, (c) causality (i.e., the doctrine of determinism), and (d) theory falsifiability. Graduate students must be trained to identify and apply these concepts within their own research beyond understanding standardized definitions and examples.

Finally, we agree with Fried when he states that “applied psychologists rarely receive training in theory building” (p. 284). Although graduate programs in psychology require students to complete research methodology courses that discuss the foundations of scientific research, most courses cover such a wide breadth of different topics that many concepts, such as theory building, are not explored in depth or with respect to application. Additionally, many methodology courses are not specific to research subfield and largely focus on broad topics related to the overall development of research standards; as a result some students may grasp the general concepts and historical progression of developing and testing theories, yet neglect how these aspects are applied within their subfield or how this knowledge should be incorporated in their own research. Perhaps to mitigate, universities should require students to complete a general research methodology course covering the basics, followed by smaller, targeted methodology courses focusing on measurement and theory-building within specific subfields.

The difficulty is that these topics are all embedded within one another and weaknesses in one area contaminate the remainder. For the sake of future research in the field of psychology, we reiterate our suggestions for improving graduate programs for the betterment of student researchers: (1) research methodology training should involve an initial global teaching of broad concepts, followed by applied training courses for subfields, and a greater focus on psychometric issues related to reliability and validity; (2) statistics courses should put greater emphasis on data-diagnostics and making appropriate decisions throughout the data-cleaning and analysis process that are justified with support; (3) the tenets of philosophy of science (replicability, transparency/verifiability, causality, and falsifiability), must be covered

more extensively and from an applied perspective, rather than being glossed over as esoteric philosophical concepts; (d) In fact, students should be required to take philosophy of science and theory-building courses to instill a greater respect for scientific inquiry, as well as rigorous experience with developing and evaluating theoretical constructs and frameworks.

References

- Baddeley, A. D., & Hitch, G. (1974). Working Memory. *Psychology of Learning and Motivation*, 8, 47–89. [http://doi.org/10.1016/S0079-7421\(08\)60452-1](http://doi.org/10.1016/S0079-7421(08)60452-1).
- Bollen, K. A., & Pearl, J. (2013). Eight myths about causality and structural equation models. In S. Morgan (Ed.), *Handbook of causal analysis for social research* (pp. 301–328). Dordrecht: Springer. [10.1007/978-94-007-6094-3_15](https://doi.org/10.1007/978-94-007-6094-3_15).
- Conway, A. R. A., Kovacs, K., Hao, H., & Snijder, J. (2019). *General intelligence explained (Away)*. Manuscript in preparation.
- Cowan, N. (2005). *Working memory capacity*. Hove, UK: Psychology Press.
- Detterman, D. K. (1987). Theoretical notions of intelligence and mental retardation. *American Journal of Mental Deficiency*, 92(1), 2–11.
- Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, 11(1), 19–23. doi:10.1111/1467-8721.00160
- Fried, E. I. (2020, February 7). Lack of theory building and testing impedes progress in the factor and network literature. <https://doi.org/10.31234/osf.io/zg84s>.
- Gergely, G., Nádasdy, Z., Csibra, G., & Bíró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, 56(2), 165–193. doi:10.1016/0010-0277(95)00661-H
- Halford, G. S., Wilson, W. H., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *The Behavioral and Brain Sciences*, 21(6), 803–865. doi:10.1017/s0140525x98001769
- Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General*, 133(2), 189–217. <http://doi.org/10.1037/0096-3445.133.2.189>
- Kievit, R. A., Romeijn, J. W., Waldorp, L. J., Wicherts, J. M., Scholte, H. S., & Borsboom, D. (2011). Mind the gap: A psychometric approach to the reduction problem. *Psychological Inquiry*, 22(2), 67–87. doi:10.1080/1047840X.2011.550181
- Kovacs, K., & Conway, A. R. (2016). Process overlap theory: A unified account of the general factor of intelligence. *Psychological Inquiry*, 27(3), 151–177. doi:10.1080/1047840X.2016.1153946
- Kovacs, K., & Conway, A. R. (2019a). What is IQ? Life beyond “general intelligence”. *Current Directions in Psychological Science*, 28(2), 189–194. doi:10.1177/0963721419827275
- Kovacs, K., & Conway, A. R. (2019b). A unified cognitive/differential approach to human intelligence: Implications for IQ testing. *Journal of Applied Research in Memory and Cognition*, 8(3), 255–272. doi:10.1016/j.jarmac.2019.05.003
- Kovacs, K., Molenaar, D., & Conway, A. R. (2019). The domain specificity of working memory is a matter of ability. *Journal of Memory and Language*, 109, 104048. doi:10.1177/0963721419827275
- Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all? *Journal of Personality and Social Psychology*, 86(1), 148–161. doi:10.1037/0022-3514.86.1.148
- McFarland, D. (2020). The effects of using partial or uncorrected correlation matrices when comparing network and latent variable models. *Journal of Intelligence*, 8(1), 7–2011. doi:10.3390/jintelligence8010007

- Meltzoff, A. N. (1988). Infant imitation after a 1-week delay: Long-term memory for novel acts and multiple stimuli. *Developmental Psychology*, 24(4), 470–476. doi:10.1037/0012-1649.24.4.470
- Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In William G. Chase (Ed.), *Visual information processing*. New York: Academic Press.
- Norman, D. A., & Shallice, T. (1980). *Attention to action. Willed and automatic control of behavior* (CHIP Report No. 99). San Diego: University of California.
- Oberauer, K. (2009). Design for a working memory. *Psychology of Learning and Motivation*, 51, 45–100.
- Schmank, C. J., Goring, S. A., Kovacs, K., & Conway, A. R. A. (2019). Psychometric network analysis of the Hungarian WAIS. *Journal of Intelligence*, 7(3), 21. doi:10.3390/jintelligence7030021
- Spearman, C. (1904). "General intelligence" objectively determined and measured. *The American Journal of Psychology*, 15(2), 201–292. doi:10.2307/1412107
- Thomson, G. H. (1916). A hierarchy without a general factor. *British Journal of Psychology*, 1904-1920, 8(3), 271–281. <http://j.2044-8295.1916.tb00133.x> doi:10.1111/j.2044-8295.1916.tb00133.x
- Van der Maas, H. L., Kan, K. J., & Borsboom, D. (2014). Intelligence is what the intelligence test measures. *Journal of Intelligence*, 2(1), 12–15. doi:10.3390/jintelligence2010012