Confirmatory Theory Testing: Moving Beyond NHST

By Keith Widaman
Past President of Division 5

Meehl’s (1967) paradox emphasizes the differences between the physical sciences, where advances in measurement make it harder to corroborate theories, and psychological science, where measurement advances, ironically, make it easier to corroborate theories. Hypotheses in the physical sciences usually specify precise numerical results, whereas in the psychological sciences, hypotheses typically specify directional results. In psychology, much theory development has relied on null-hypothesis significance testing (NHST).

NHST epitomizes an exploratory approach to theory corroboration. NHST is somewhat counter-intuitive on its face, because rather than testing the prediction of the theory, it sets up a “straw man” representing the opposite of the theory’s prediction. For instance, if our theory predicts that a 10-week trial of cognitive-behavioral therapy reduces depression symptoms, NHST tests the hypothesis that the therapy has no effect.

Confirmatory strategies offer a resolution to Meehl’s paradox, because they offer a generalized approach to testing substantive (non-null) hypotheses. These approaches can be relatively simple to implement and interpret, using regression methods that provide straightforward interpretations of beta weights. In addition, confirmatory approaches permit the use of fixed, free, and constrained parameters—as commonly used in structural equation modeling—so they can support the testing of models specified in many different ways. Most importantly, and in contrast to NHST, confirmatory strategies allow the hypothesized model to be embodied in the hypothesis being tested.

The benefits of the confirmatory approach may become more apparent when we contrast competing hypotheses in a particular research application. Consider the study of influences of genetic and environmental factors on risk for negative outcomes such as depression. Research on gene X environment (GxE) interactions often investigates GxE interactions under one or both of two competing models of depression risk. First, the diathesis-stress model (Caspi et al., 2003; Belsky & Pluess, 2009) predicts that a problematic genetic allele creates a diathesis, or vulnerability, to depression. Expression of the depressive phenotype, however, depends on stress. In the absence of stress, there are no differences in depressive symptoms between the high-risk group (genetic diathesis present) and the low-risk group (genetic diathesis

(Continued on page 3)
The Score is the official newsletter of APA Division 5—Division for Quantitative and Qualitative Methods—and is published quarterly in January, April, July and October. In keeping with this mission, The Score publishes the division’s business meeting minutes, committee reports, and announcements.

In addition, where appropriate and space permits, short articles (800–1000 words) on technical issues and professional activities of Division 5 members, or on topics of current interest may be accepted. Brief announcements and calls for presentations related to conferences or meetings of particular interest to Division 5 members may also qualify. Submissions should be sent to The Score Editor, David Herzberg: dherzberg@wpspublish.com.

Submission deadlines are one month prior to publication: March 1 for the April issue, June 1 for July, September 1 for October, and December 1 for January.

The Score is published solely online and distributed via e-mail notification. Division 5 members receive the e-mail notice through the Division 5 DIV5ANN email listserv (see the box below).

Guidelines for advertising appear elsewhere in this issue. Paid advertisements are solicited from a variety of sources and are not officially endorsed by Division 5.

Guidelines for the “What’s New?” column are provided with the column.

Urgent announcements should be submitted to the Division 5 e-mail lists, described in the box below.

---

**E-mail Lists**

Keep up with the latest Division 5 news through its two e-mail listservs.

**DIV5** serves as a vehicle for discussion among members on topics related to evaluation, measurement, statistics, assessment, and qualitative inquiry.

**DIV5ANN** is used exclusively for announcements from Division leadership regarding issues such as elections, calls for nomination to boards and committees, the convention program, or policy changes. This is a “one-way” listserv that does not support listwide replies (that is, it is not structured to support discussion). All members have been added to the listserv, and new members are added as they join. We strongly encourage all members to remain on the listserv in order to receive the newsletter and other important division news. We promise to keep the number of messages to a minimum.

To subscribe to the discussion list, div5@lists.apa.org, send a message to LISTSERV@LISTS.APA.ORG. Leave the Subject line blank. In the body of the e-mail, include the following:

**SUBSCRIBE DIV5**

You can manage your listservs via http://listsserve.apa.org. If you have any questions, contact Mark Daniel at Mark.Daniel@pearson.com.
Confirmation Theory Testing: Moving Beyond NHST

(Continued from page 1)

absent). Group differences emerge only when activated by stress.

The confirmatory approach provides a basis for evaluating two versions of this model. The strong diathesis-stress model, shown in Figure 1A, suggests that the low-risk group (solid line) will be completely unaffected by stress loading. The weak diathesis-stress model, shown in Figure 1B, proposes that the low-risk group will be affected (will show symptoms of depression) as a function of stress, but the effect will be smaller than that with the high-risk group.

A competing theory of depression vulnerability is the differential-susceptibility model (Belsky & Pluess, 2009). This model allows for the high-risk group to have better outcomes than the low-risk group under certain environmental conditions. In other words, the genetically vulnerable group may show more depression than the low-risk group when stress is high, and less depression than the low-risk group when stress is low. Thus, the differential-susceptibility model predicts a “crossover” interaction effect.

As with the diathesis-stress model, there are strong and weak versions of the differential-susceptibility model. The strong version, shown in Figure 1C, holds that the low-risk group is unaffected by changes in the environment, whereas the weak version, shown in Figure 1D, asserts that the low-risk group is affected by the

Figure 1: Predicted outcomes under the diathesis-stress and differential susceptibility models
environment, but to a lesser degree than the high-risk group.

The competing predictions in Figures 1A through 1D present the two versions of each model, illustrating two key differences between the models. First, the two models differ substantively, a difference represented by the position of the crossover point on the graph. Second, the two models differ quantitatively, in the number of parameter estimates required to test hypotheses generated by the models.

A primary advantage of the confirmatory approach now becomes apparent. It turns out that the weak differential-susceptibility hypothesis can be specified as a generalized, four-parameter model, of which the other three competing hypotheses are special cases. To illustrate, let us first examine the usual regression approach to testing models of genetic vulnerability.

Typically, such models test for a main effect of environment, \( X \), a main effect of genes, \( G \), and the interaction of genes and environment \( (G \cdot X) \), using an equation of this type:

\[
Y = B_0 + B_1 X + B_2 G + B_3 (G \cdot X) + E
\]

where \( B_i \) is the intercept for non-risk group when \( X = 0 \), \( B_1 \) is the effect of environment \( X \) for the non-risk group, \( B_2 \) is the intercept difference for the risk group when \( X = 0 \), and \( B_3 \) is the slope difference for the high-risk group. In this approach, the interaction term \( (B_3) \) must be significant for the analysis to proceed.

To move beyond the usual regression approach, it is necessary to incorporate the substantive differences between the competing models. As mentioned earlier, these substantive differences are represented by the crossover point, which can be estimated as \( C = -B_2/B_3 \) (see Aiken & West, 1991).

This method of estimating \( C \), however, yields neither a standard error term nor a confidence interval, and so the interaction term (product of genes and environment terms) must still be significant in order to proceed.

Estimating the crossover point does permit a new formulation of the regression equation, one where \( X \) is centered at the crossover point, \( C \). As before, the model includes a main effect of environment, \( X \), a main effect of genes, \( G \), and the interaction of genes and environment, which is now represented as \( (G \cdot (X - C)) \):

\[
Y = B_0 + B_1 (X - C) + B_3 (G \cdot (X - C)) + E
\]
where $B^*_0$ is the predicted value of $Y$ at the crossover point, $B^*_1$ is the effect of environment $X$ for the non-risk group, $C$ is the estimate of the crossover point, and $B^*_3$ is the slope difference for the high-risk group.

When $X$ is centered at the crossover point, in the manner shown above, the resulting equation represents the weak differential susceptibility model. To see why this model subsumes the other three competing models, consider again the graphs in Figure 1. Visual inspection suggests that the weak differential susceptibility model is the least constrained of the three models. It is less constrained than the strong versions of both models, because it permits the low risk group to be affected by the environment, reflected in the $B^*_1$ parameter in the re-parameterized model. And, it’s less constrained than the weak diathesis stress model, because it allows that, under certain optimal environmental conditions, the high-risk group may have a better outcome than the low-risk group.

Figure 2 provides another perspective on the hierarchical nature of the re-parameterized models. The weak differential susceptibility model requires the estimation of four parameters: $B^*_0$, $B^*_1$, $B^*_3$, and $C$. Each of the other three models requires estimation of only a subset of these four. Solving the regression equation for the weak differential susceptibility hypothesis, therefore, simultaneously provides a solution that is relevant when evaluating the other three regression equations under study. This method epitomizes a confirmatory approach, therefore, because the four competing models are embodied in the hypothesis test, and the hypothesis that “nothing is going on” is nowhere to be found. As an answer to Meehl’s paradox, the confirmatory approach stands in contrast to NHST, where the only hurdle a theory needs to surmount is to predict an effect that is greater than nothing. And, the confirmatory approach is quite adaptable, applicable across most or all domains of psychology, if only we can embody our predictions in clearly specified regression models.

References


Greetings, graduate students of Division 5! My name is Veronica Cole, and I am the representative to the American Psychological Association of Graduate Students (APAGS) for Division 5. As a fourth-year graduate student in UNC Chapel Hill’s Quantitative Psychology program working with Dr. Daniel Bauer, most of my research thus far has focused on improving measurement in commonly used quantitative methods such as mixture models, multilevel models, and integrative data analysis (IDA). As I suspect is the case with students in many other graduate programs, I consider myself so privileged to be surrounded by great faculty and peers, and I try to get the most out of the tremendous opportunities my education affords me. It is my hope that we can make use of resources within Division 5 and APAGS to help every graduate student within the division get the most of their graduate school experience. Admittedly, until recently I was unaware of a great number of the resources Division 5 and APAGS make available to the graduate student community, so I would like to take the opportunity to describe the resources that are available, and how students may get involved.

APAGS is the main committee of graduate students within APA, and serves as a voice for our concerns and interests, both within the larger organization of APA, and to the world at large. With representatives from almost all divisions of APA, APAGS provides a forum for graduate students within each division to communicate with one another and learn what’s going on in different areas of the field. Additionally, APAGS provides resources, advocacy opportunities, and access to a number of APAGS-specific funding opportunities. Visiting APAGS’ webpage (http://www.apa.org/apags/resources/index.aspx) reveals the impressive breadth of resources APAGS has assembled for the graduate student community—from gradPSYCH, the magazine for psychology graduate students, to lists of resources for graduate students with disabilities, to guides to applying to and paying for graduate school.

Representing Division 5 to the APAGS Division Student Representative Network (DRSN) has been a great learning experience so far, as I have gotten the chance to hear from other divisions’ representatives about the unique aspects of their fields, as well as more general concerns and experiences we all share. Graduate students seeking further involvement may run for any one of APAGS’ leadership positions. However, it is important to note that any graduate students who are members of APA are already members of APAGS, so even those who are not interested in a leadership role may take advantage of the benefits of APAGS membership. Visit the webpage linked above or subscribe to the APAGS listserv for updates!

Additionally, there are a number of resources within Division 5 itself that are geared toward the graduate student population. The division provides a great deal of programming at the yearly APA convention that is of particular interest to graduate students—from a poster session where graduate student work may be presented, to symposia, to social and networking hours. We are also working on providing more organized opportunities for graduate students within the division to communicate with one another outside of the APA convention itself, particularly by increasing the Division 5 graduate student presence on social networks such as Facebook and Twitter. Be on the lookout for updates about these efforts, and please contact me if you have any interest or suggestions!

Involvement in APA, both through APAGS and Division 5, has given me the sense that graduate students within Division 5 are in an especially strong position to advocate for the graduate student community. As quantitative and qualitative psychologists, one of our critical goals is the measurement and operationalization of important aspects of the human experience. In many ways, this is one of the most important elements of advocacy: After all, in order to advocate for any cause, one must collect and analyze a great deal of data, formally or informally, in order to fully understand the needs of the people for whom one is advocating. Who better to collect and analyze these data, and connect the data to an agenda for advocacy, than the graduate students themselves?

Furthermore, given that a great number of us are methodologists, we likely get a greater-than-average number of opportunities to interface with a wide breadth of other graduate students. I know that, in my role as a statistics teaching assistant and research assistant in my department, I am afforded the chance to work with other graduate students in clinical, cognitive, developmental, and social psychology on a regular basis. Through this work I’ve been able to hear about their concerns, interests, and goals; this happens naturally as a function of methodological work itself, through which I get the chance to learn about—and, hopefully, help with—other students’ research interests. These other graduate students’ experiences are both similar to and different from my own, and I might not have gotten the chance to hear about these experiences if we hadn’t been brought together by statistics class or research—and I imagine that a number of other Division 5 graduate students interface with a wide range of other graduate students in much the same way.

The main question that follows from these observations is: How do we leverage our role as quantitative and qualitative psychologists to be a voice for the graduate student community? I have a few...
The Role of Assessment in the Clinical Psychology Internship

By Colleen Reichmann

Predoctoral internship year is an incredible time in a student’s life. Graduate classwork is a thing of the past, and the idea of having a career as a licensed psychologist no longer seems like some sort of a pipe dream. I myself am currently about three months into my internship year. My experience thus far has been full to the brim with providing therapy, participating in journal clubs and attending didactic trainings. I have also made it a point to utilize supervision to the utmost extent. I am well aware that this is the last year of intensive supervision that I will have; hence I have a strong desire to take in as much knowledge and experience of my supervisors as possible. As a Student Editor for The Score, I am always looking for opportunities to incorporate assessment into my internship.

My experience thus far involves providing group and individual therapy to individuals suffering from eating disorders in an acute care inpatient eating disorders hospital. In just three months, I have grown more in my knowledge about eating disorders and acquired more skills as a clinician than I would have ever thought possible. The vast majority of my time has been spent providing group therapy. I have been leading and co-leading an array of cognitive behavioral psychoeducational groups, dialectical behavioral therapy groups and psychotherapy process groups. Additionally, while assessment is not the emphasis of my internship experience, it certainly plays a role. I have a specified assessment supervisor whom I meet with bimonthly, along with the other predoctoral interns.

The actual assessment experience that I have had has been eye-opening. For one, I was rather surprised to find that projective testing is routinely included in the assessment batteries conducted at my site. My training and research has been focused on empirically validated, behaviorally oriented assessment techniques; hence I have had little training with projective measures. In this sense, I have found supervision to be helpful to my understanding of these unique assessment tools. I have received training on how the information yielded by projective testing can be combined with other assessment information to produce a truly multifaceted case conceptualization. I am now eager to undergo more Rorschach training, so that I will be able to administer this test and integrate these findings into other psychological testing data. I know that the Rorschach is controversial, and that its validity has been questioned, and I want to gain the knowledge necessary to form my own opinion about the utility of this test.

I have just recently conducted my first assessment battery, and have found the information that it yielded to be highly useful for my clinical interpretations of the patient with whom I was working. One important fact that I have learned about assessment among individuals with eating disorders is that there are often cognitive impairments that occur as a side effect of malnutrition. Hence when patients are displaying delayed processing speed or impaired focus, the root of these issues becomes rather difficult to decipher.

I was recently asked to utilize cognitive and personality testing in order to rule out cognitive impairment for a 25-year old female patient who was experiencing delayed processing speed (Note: background information on this case has been changed to protect confidentiality). This patient had been diagnosed with anorexia two years ago, and observers had described a steady cognitive decline over the past year. I administered the Weschler Abbreviated Scale of Intelligence (WASI; Weschler, 1999) and the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS; Randolph, Tierney, Mohr, & Chase, 1998). I also administered the Eating Disorder Inventory-3 (EDI-3; Garner, 2004) and the Millon Clinical Multiaxial Inventory 3 (MCMI-3; Millon, Millon, Davis, & Grossman, 2009). I chose the WASI and the RBANS to clarify the nature of the cognitive impairment. The EDI-3 was utilized in
order to elucidate the presence and intensity of eating disordered thoughts and beliefs. Finally, the MCMI was employed to clarify whether personality attributes or other psychopathology, such as depression, may be contributing to the patient’s cognitive presentation.

This battery yielded information that was invaluable to both diagnosis and discharge planning. The results of the WASI suggested that this individual scored in the average range on verbal reasoning, and in the low average range on perceptual reasoning. The RBANS, however, indicated mild impairment in both immediate and delayed memory, and visual spatial abilities. Personality testing did not reveal any Axis II diagnoses; however, a general tendency towards dysthymia and a pessimistic outlook was noted. Interestingly, the EDI-3 indicated a mild level of eating disordered pathology, which suggested a lack of judgment and insight. This assessment battery helped me to gain a deeper comprehension of what psychological, neurological, and personality variables may be interacting and impacting this individual’s diagnosis and presentation. I was able to suggest that her observed memory impairment and visual-spatial abilities were most likely due to malnutrition, because research has demonstrated that these are the cognitive domains that starvation typically affects. Additionally, I noted that dysthymia and low insight were likely contributing factors to her eating disorder.

The value of assessment for psychological practice is now quite clear to me. Case conceptualization is enhanced by testing, because it yields a much more comprehensive understanding of clients’ problems than a simple clinical interview. Specifically, it adds quantitative data to the largely qualitative information gained from the interview. I am grateful to have the experience of conducting these assessment batteries while on internship, because I will continue to have the opportunity to gain a deeper understanding of how to both administer tests and interpret their results with an eating-disordered population.

References
Have you published a new psychological test or testing product; a book on advanced statistics, measurement, or evaluation; an interesting website or other Internet group related to measurement, statistics, or evaluation; or a computer program useful to Division 5 membership? If so, we would like to include an announcement of about 100 words in this column. We would also appreciate any suggestions, or feedback, on how this section of the newsletter can better serve the Division 5 membership.

Please take the opportunity to share information with colleagues through your contributions to this column. Please send announcements and/or product literature to Associate Editor Michael Edwards: edwards.134@osu.edu

**What's NEW...**

**Health Measurement Scales: A Practical Guide to Their Development and Use (Fifth Edition)**
By David L. Streiner, Geoffrey R. Norman, and John Cairney
Published in November 2014 by Oxford University Press ($57.50 paperback)

This is the fifth edition of a book about the construction of scales used by health professionals. Chapters follow the sequence researchers would use: developing items; choosing scoring formats; assessing the items; looking for various biases that can affect responses; checking reliability and validity (with an extended discussion of generalizability theory); and considering various methods of administration. Chapters also discuss item response theory, measurement of change, ethical issues, and how to write articles about findings. All chapters have been updated to reflect the latest research in the area, and a flow chart guides readers through the various stages of scale development and where to find that information in the book. Useful for students, clinicians, and experienced researchers.

**Bayesian Statistics for the Social Sciences**
By David Kaplan
Published in July 2014 by Guilford Press ($51 hardback; $51 e-book)

Bridging the gap between traditional classical statistics and a Bayesian approach, this book provides readers with the concepts and practical skills they need to apply Bayesian methodologies to their data analysis problems. Part I addresses the elements of Bayesian inference, including exchangeability, likelihood, prior/posterior distributions, and the Bayesian central limit theorem. Part II covers Bayesian hypothesis testing, model building, and linear regression analysis, carefully explaining the differences between the Bayesian and frequentist approaches. Part III extends Bayesian statistics to multilevel modeling and modeling for continuous and categorical latent variables. Kaplan closes with a discussion of philosophical issues and argues for an “evidence-based” framework for the practice of Bayesian statistics.

**A Survivor's Guide to R: An Introduction for the Uninitiated and the Unnerved**
By Kurt T. Gaubatz
Published in May 2014 by Sage ($35 paperback)

Focusing on developing practical R skills rather than teaching pure statistics, A Survivor’s Guide to R provides a gentle yet thorough introduction to R. The book is structured around critical R tasks, and focuses on applied knowledge, rather than abstract concepts. The easy-to-read approach helps students with little or no background in statistics or programming to develop real-world R skills through straightforward coverage of R objects and functions. Focusing on real-world data, the challenges of dataset construction, and the use of R’s powerful graphing tools, the guide is written in an accessible, sympathetic, even humorous style that ensures students acquire functional R skills they can use in their own projects and carry into their work beyond the classroom.

**Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory (Fourth Edition)**
By Juliet Corbin and Amselm Strauss
Published in December 2014 by Sage ($72 paperback)

Offering eminently practical advice, Basics of Qualitative Research presents methods that enable researchers to analyze, interpret, and make sense of their data, and ultimately build theory from it. The authors walk readers step-by-step through the research process—from the formation of the research question through several approaches to coding, analysis, and reporting. Packed with definitions and illustrative examples, this accessible book concludes with chapters that present criteria for evaluating a study, as well as responses to common questions posed by students of qualitative research. New end-of-chapter “Insider Insights” contributed by qualitative researchers give readers a sense of what it’s like to work in the field. Significantly revised, this new edition remains a foundational volume in the study of qualitative methods.
APAGS, Division 5, and Our Graduate Student Community

(Continued from page 6)

ideas, and all of them would benefit from the involvement of any graduate student who is interested in helping.

One of the main things that I would like to do is a data collection effort about Division 5 programming; in particular, I am interested in hearing from other graduate students what sorts of topics they would be interested in learning about, both at the APA convention and online. The survey will be administered in the upcoming semester—and if you have any interest in helping to design or distribute this survey, please do not hesitate to let me know.

Additionally, becoming involved in the broader organizational structure of APAGS is a great way to serve as an advocate for fellow graduate students, both within the division and outside of it. Though I serve as the division’s representative to the DRSN, this is just one area of APAGS—there are a number of broader committees and sub-committees to which one can be elected. See the APAGS governance page (http://www.apa.org/apags/governance/index.aspx) for more details.

Again, I am really grateful for the chance to represent the interests and concerns of Division 5 graduate students to APAGS-DRSN. Please email me (vcole@email.unc.edu) with your thoughts, concerns, and suggestions!

Advertise in The Score

*The Score* is the newsletter of the American Psychological Association’s Division 5—Division for Quantitative and Qualitative Methods. Division 5 is concerned with promoting high standards in both research and practical application of psychological assessment, evaluation, measurement, and statistics. Approximately half of the Division 5 members are university faculty members in quantitative psychology, psychometrics, educational psychology, or industrial-organizational psychology and half are engaged in careers in industry, including the areas of individual and large-scale assessment. More than 1,000 Division 5 members receive *The Score* each quarter.

Advertisements in *The Score* may be in the form of display advertisements or job announcements. Both types of ads can include graphics and other design features and can be submitted as text or camera-ready display art. Prices for advertisements and size requirements are provided in the accompanying table. Submission deadlines are 45 days prior to publication: February 15 for the April issue, May 15 for July, August 15 for October, and November 15 for January. To advertise in *The Score*, please contact Editor David Herzberg at dherzberg@wpspublish.com.

<table>
<thead>
<tr>
<th>Size</th>
<th>Dimensions</th>
<th>Display Ad Price</th>
<th>Job Announcement Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full page</td>
<td>7.125&quot; × 9.5&quot;</td>
<td>$235</td>
<td>Not available</td>
</tr>
<tr>
<td>Half page</td>
<td>7.125&quot; × 4.75&quot;</td>
<td>$175</td>
<td>$120</td>
</tr>
<tr>
<td>Third page</td>
<td>4.75&quot; × 4.75&quot;</td>
<td>$125</td>
<td>$90</td>
</tr>
<tr>
<td>Sixth page</td>
<td>2.375&quot; × 9.5&quot;</td>
<td>$90</td>
<td>Free/$55</td>
</tr>
<tr>
<td></td>
<td>2.375&quot; × 4.75&quot;</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: Insertion orders for four consecutive issues receive a 15% discount. First sixth page job ad each year free, thereafter $55.