

The Score



DIVISION 5 PERIODICAL VOLUME 45 ISSUE 4

EDITOR: GILL SITARENIOS

ASSISTANT EDITOR: JOANNA SOLOMON



In this issue . . .

FREDERICK WERTZ 2

Our Challenge: Heterogeneity and Unity in Psychological Methodology

KEVIN WILLIAMS 5

Measuring Bias in Job Performance Evaluation: Applying Workplace Contextualization and

DEBORAH BANDALOS 9

Research on culturally relevant items in educational assessments

JAMES CHRISTOPHER HEAD 14

Division 5 Awards 2023

ALSO IN THIS ISSUE 17

What's New in Books (p. 17), Publication Corner (p. 19), Links to previous issues (p. 21),
Division 5 Membership Information (p. 22); Advertising in The Score (p. 23).



Frederick Westz

DIVISION 5 PRESIDENT MESSAGE -- OUR CHALLENGE: HETEROGENEITY AND UNITY IN PSYCHOLOGICAL METHODOLOGY

Frederick Westz

The vitality and generativity of APA's Division of Quantitative and Qualitative Methods were evident at the recent APA Convention in Washington, DC. The research and persons on which our Division bestowed awards could not be more diverse and impressive. Assuming the position of President of the Division was a special pleasure for me given our newly established tradition to deliberately engage the Presidential Trio in planning initiatives. Alex Beaujean's (past-past-President) influence continues as Debbi Bandalos (Past President), Rachel Fouladi (President-Elect), and I continue to increase communication and collaboration with the Division's Executive Committee. There is excitement about the continuation of our Mentor Program, a new Salon Series, the prospect of Division publication on psychological methodology, a possible grant or award in the area of mixed methods, and next year's convention theme: *Nurturing Roots, Branching Innovation: Cultivating Growth and Synthesis in Psychological Methods and Methodology*.

Our vision and initiatives are responding to Alex Beaujean's call, in last year's presidential address ("Going Back to Move Forward: A Return to Methodology"), for renewed emphasis on methodology in Division 5, APA, and psychology more generally. Alex urged us to rise above the production of research and concerns and the technical details of methods to reflect more broadly and critically on our methodological *norms*, that is, the assumptions, rationale, goals, evidence, limits, interrelations, and cultural-historical contexts of psychological methods. We can overcome the silos in areas of evaluation, measurement, statistics, assessment, and qualitative inquiry and among the various content areas of psychology with normative perspective--how psychological phenomena *ought* to be investigated.

We look forward to dialogue and collaboration among our Sections (EMS, Assessment, and SQIP), with other APA divisions, and even other disciplines. The emerging emphasis on how our knowledge *ought* to be acquired, based on ontological, epistemological, and axiological considerations, was featured in such August convention symposia as "The Value and Applications of Mixed Methods in Psychological Research"; "Sister Disciplines in Dialogue: The Virtues and Necessity of Interdisciplinary Exchange"; "Positivist Methods and the Decontextualization of Race: A Critical Discourse Analysis"; "Combining Focus Groups with Surveys or Interviews to Enhance Cultural Validity"; "Suppressing Subjectivity: The Costs of Semantic Drift in 'Evidence-Based Practice in Psychology'"; and "Approaching Significance: How to Navigate the Wild World of Statistics" as well as presentations by Chris Green on "The Rhetorical Force of Statistical Language in Psychology" (co-winner of Division 1's Arthur W. Staats Award Lecture for Unifying Psychology) and president Debbi Bandalos's, "Culturally Responsive Testing: Perspectives from Education and Psychology."

Our Division's initiatives are works in progress that welcome your input. A new Salon Series is being spearheaded by Debbi Bandalos. We are working on one Salon focusing on the use of qualitative methods by psychologists and philosophers of science to investigate the practices of natural scientists and another on

methodological and social implications of indigenous psychological research in Native American communities. We hope you will suggest additional ideas and possible presenters for this year's salon series--please feel free to email us.

Our Program Co-Chairs for the 2024 APA Convention in Seattle are Gabriel Velez and Joshua Langfus, who represent qualitative and quantitative methods respectively. We especially welcome submissions of the following kinds:

- Broad and critical reviews of psychological methodology as we approach the quarter century
- Cross-cutting presentations that span, compare, and integrate the full spectrum of psychological methods
- Critical studies of the values, goals, rationale, limits, and relationships (e.g., incongruities, complementarities) of existing methods
- The implications for methods of the nature and diversity of psychological subject matter
- How methods and methodology can better bridge the gulf between scientific research and practical applications of knowledge, including transformative social change.
- Psychology's relations with other (natural and social) sciences and investigative approaches in the humanities and arts

My own long-standing interest in these difficult issues goes back to my undergraduate studies of psychology, which motivated me to probe the methodological foundations of our science. I was fascinated by our history, especially the separation from philosophy of science, defined by its empirical methods parallel to if not adopted from physical sciences. I appreciated the creativity, ingenuity, and precise rationality of measurement, statistics and controlled experiment yet was also preoccupied with questions about the nature of psychological life that spans conscious, behavioral, and psycho-physiological, personal phenomena. I was struck by the criticism, debate and controversy that extended from the founding of psychology through the 20th century. My questions led me to the work of mathematician, philosopher of science, and founder of phenomenology Edmund Husserl, his teacher Franz Brentano, kindred thinker Wilhelm Dilthey, and the related European movements in Gestalt, personalistic, and existential approaches that utilized qualitative methods. However, in the 1970s, the term "qualitative" had not even been coined, and little formal education and knowledge of such methods was available.

I began to clarify the methods and methodological norms of qualitative psychologists with the emerging phenomenological psychological methods to study lived experience developed by Amedeo Giorgi. I found similar methods used throughout the history of psychology in the work of such pioneers as Wundt, James, Freud, Maslow, Kohlberg, and virtually all psychologists without formal specification or explicit methodological accounts. After living through and documenting the formal emergence of qualitative methods and methodologies during the 1970s and 1980s, I am surprised by the evolving legitimation and integration of these methods that was evident in APA Division 5's welcoming inclusion of qualitative psychologists and its name change to the "Division of Quantitative and Qualitative Methods." My comparative methodological studies of qualitative methods have convinced me of fundamental commonalities in the heterogeneity of qualitative methods and their inextricable relatedness to all quantitative methods in psychology. Given innovations in all areas of psychology and growing plurality among our methods, we need a foundational,

overarching understanding of the purposes, limits, and interrelations of our methods that informs us of how valid knowledge of psychological phenomena is best acquired.

Now is the time for quantitative and qualitative psychologists to come together, learn openly from and with each other. What we do often fruitfully exceeds our formal knowledge of method on the one hand and yet too often falls short of what we ought to do on the other hand. My worry is that our philosophies of psychology, without good education or careful study, have become caricatures of outdated views that divide and mislead us rather than provide a solid foundation that reflects the diverse best practices of Division 5 members and the variety of psychologists. We have the necessary expertise to engage in fresh, critical thinking in the new frontiers of 21st century psychology and are positioned to meet the challenge of forging better, increasingly integrated, and more broadly shared methodological knowledge. I hope you look forward, as I do, to learning about unknown and barely known methods from our colleagues in and beyond Division 5 in the service of larger understandings of psychological methodology.

*Kevin Williams*

MEASURING BIAS IN JOB PERFORMANCE EVALUATION: APPLYING WORKPLACE CONTEXTUALIZATION AND EMPIRICAL ASSESSMENT DEVELOPMENT CRITERIA

Kevin Williams

Author Note: The author is part of the Educational Testing Service (ETS)

The ubiquity and consequences of job performance evaluations (Campbell & Wiernik, 2015) necessitate accurate responding in both self-report and other-report assessments. However, response distortion or response bias – described as “a systematic tendency to respond to a range of questionnaire items on some basis other than the specific item content” (Paulhus, 1991; p. 17) – represents a significant threat to the veracity of performance evaluations. Examples of response bias that may occur in either self- or other-reports include positive impression management (PIM; i.e., socially desirable responding, positive bias, “faking good”) and negative impression management (NIM; i.e., negative bias or “faking bad”; see reviews by McGrath et al., 2010; Paulhus, 2002). Self-reported bias is often related to pervasive dispositional factors or temporary situational pressures (e.g., job interviews), whereas other-report bias (e.g., evaluations from supervisors or coworkers) may be influenced by ratees’ impression management (IM) tactics (Peck & Levashina, 2017), halo effects (McGill et al., 2011), raters’ self-interest (Murphy, 2008; Murphy & Cleveland, 1995), to motivate ratees (Murphy, 2008; Murphy & Cleveland, 1995), or unrepresentative observations (Berry et al., 2012; Murphy, 2008).

To detect response distortion, researchers and practitioners frequently rely on assessments called bias indicators. These indicators are usually developed for general self-report research purposes beyond the workplace (e.g., Crowne & Marlowe, 1960; Paulhus, 1998; see review by McGrath et al., 2010). Although many authors support the use of bias indicators in research and applied workplace settings (e.g., McGrath et al., 2010), meta-analyses and reviews suggest only weak empirical evidence for response distortion’s threat to validity in workplace contexts (Li & Bagger, 2006; Ones et al., 1996). Plausible explanations for these results include the questionable validity of bias indicators themselves as well as criterion coarseness (poor alignment between predictors’ and outcomes’ respective scopes; McGrath et al., 2010). That is, bias indicators may lack appropriate relevance or specificity, given that most bias indicators utilized in workplace research include noncontextualized, generic statements (Blanch et al., 2009; Goffin & Christiansen, 2003; Li & Bagger, 2006; McGrath et al., 2010).

We aimed to address these concerns by developing a response bias assessment contextualized to workplace settings. Converting non-contextualized statements (e.g., “I am a responsible person”) to workplace-contextualized statements (e.g., “I am a responsible employee”) may detect workplace response bias more accurately. Meta-analyses and reviews suggest that contextualizing assessments increases their predictive validity (e.g., Heggstad & Gordon, 2008; Shaffer & Postlethwaite, 2012). Although instruments such as

personality assessments have been contextualized in this manner (see Heggstad & Gordon, 2008; Shaffer & Postlethwaite, 2012), it appears that contextualization has yet to be applied to bias indicators.

Assessment Design and Item Pool Development. We developed a workplace-specific assessment – called the Occupational Performance Assessment – Response Distortion (OPerA-RD) scale – designed to measure biased performance ratings. The OPerA-RD includes four key features. First, the instrument exclusively includes items contextualized to the workplace. As a first step, we utilized tactics common in other impression management measures, including writing statements that describe the denial of common flaws or claiming of uncommon virtues (e.g., “I never swear”; “I always know why I like things”, Paulhus, 1991; see also McGrath et al., 2010). We then integrated workplace-relevant terminology, resulting in items such as “never misses deadlines” and “is the perfect employee”. Thus, items represent unrealistically high or low standards of performance. Second, the instrument is amenable to self- and other-reporting. Specifically, items describe behavioral or otherwise easily evaluative statements as opposed to more covert actions (e.g., thoughts, attitudes). Third, the assessment should be responsive to either over- or underreporting of workplace performance (i.e., PIM or NIM, respectively). Finally, items were selected based on their ability to detect self- and other-reported over-reporting and under-reporting of the ratee’s performance. This utilization of empirical data in the item selection process is distinctive from other IM assessments, which tend to use nonempirical (e.g., conceptual) criteria. After applying these criteria, an initial pool of 56 items was developed to detect over- or under-reporting of job performance from either a self- or other-report perspective. Respondents rate their level of agreement with these items on a Likert-type scale (i.e., 1 = strongly disagree to 5 = strongly agree). We then collected data from two independent samples for item selection (Sample 1) and initial validation (Sample 2).

Item Selection. The item selection sample included 600 currently employed individuals recruited online through Amazon Mechanical Turk (MTurk). Participants were randomly divided into six approximately equal-sized groups using a 2 (self-report vs. other-report) x 3 (fake good, fake bad, control) between-subjects design. Individuals in the other-report conditions were screened to ensure management experience, experience formally rating coworkers’ job performance, or both. Participants were instructed to respond to the OPerA-RD items as part of a simulated job performance evaluation exercise, with more specific directions provided for each group (Williams, 2022). Detailed faking instructions were provided to simulate realistic workplace scenarios in which an employee may rate themselves or a co-worker more positively or negatively than usual. Specifically, self-report fake good instructions requested participants to rate themselves more positively than usual so they could receive a raise, whereas self-report fake bad instructions requested participants to rate themselves more negatively than usual so they would receive an all-expenses-paid training opportunity (see Williams, 2022). For other-report, participants were provided motivation to fake good or bad such that the ratee would be more likely to gain a desired outcome, similar to the self-report condition. Control conditions instructed participants to rate themselves (self-report) or a coworker (other-report) accurately. All items were analyzed, and the 20 best-performing items from this first phase were tested again in a new sample.

Sample 2 included 1,318 currently employed participants recruited through MTurk. Participants were randomly assigned to one of six groups using a research design that incorporated within-subjects and between-subjects features. Two of the groups responded to the final twenty-item OPerA-RD self-report scale, once under control instructions and once under either fake good (Group 1) or fake bad (Group 2) instructions

similar to those in Sample 1 (see Williams, 2022). The remaining four groups were asked to assume the perspective of a supervisor and use the final twenty-item OPerA-RD other-report scale to rate a coworker (subordinate) whose performance they were highly familiar with and had worked with for at least one month. Participants provided co-worker ratings once under control instructions and once under either fake good (Groups 3 and 4) or fake bad (Groups 5 and 6) instructions, again similar to Sample 1. Twenty standard job performance items were embedded within the OPerA-RD to further disguise the latter items.

Within each group, the difference between mean OPerA-RD total scores under control and faking conditions was used to potentially confirm the OPerA-RD's validity as an instrument capable of detecting the over- or under-reporting of self- or other-reported job performance. Results supported the OPerA-RD's ability to detect faking in all conditions, as reflected by at least a large effect size (i.e., $d \geq |0.80|$; Cohen, 1988) in the respective expected direction. Absolute d values ranged from 1.34 to 2.28, and none of the 95% confidence intervals included values below $|0.80|$.

Discussion. Given the debatable empirical support for the threat of response distortion in workplace performance assessment, many researchers and practitioners may be convinced that these types of screening procedures should be abandoned. However, most researchers maintain their support for the use of bias detection assessments in workplace contexts such as hiring, performance evaluation, and research (Goffin & Christiansen, 2003). Ideally, such an assessment should be able to detect over- or underreporting of job performance by oneself or others, meaning items should reference behaviors or attributes that are observable or otherwise amenable to external evaluation. To this end, the OPerA-RD – developed using a distinctively contextualized and empirical methodology – may advance both science and practice. More generally, our results may support the assessment design and item selection processes we utilized for future scale development. Additionally, this general methodology could be evaluated using alternative statistical criteria for selecting items (e.g., item response theory).

In practical applications, as with any high-stakes setting, basing decisions on the results of a single assessment is not recommended as legitimate, transient factors (e.g., applicant mood, stress, distractions) may contribute to construct-irrelevant variance. Instead, extreme OPerA-RD scores may warrant further scrutiny of an applicant (e.g., through interviews or other additional assessments) to support initial results. Additional details regarding this study are provided in Williams (2022). The development of this assessment may provide a useful tool for future research and applied settings alike.

References

- Berry, C.M., Carpenter, N.C., & Barratt, C.L. (2012). Do other-reports of counterproductive work behavior provide an incremental contribution over self-reports? A meta-analytic comparison. *Journal of Applied Psychology, 97*, 613-636.
- Blanch, A., Aluja, A., Gallart, S., & Dolcet, J.-M. (2009). A review on the use of NEO-PI-R validity scales in normative, job selection, and clinical samples. *European Journal of Psychiatry, 23*, 121-129.
- Campbell, J.P., & Wiernik, B.M. (2015). The modeling and assessment of work performance. *Annual Review of Organizational Psychology and Organizational Behavior, 2*, 47-74.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Crowne, D.P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*, 349- 354.

- Goffin, R.D., & Christiansen, N.D. (2003). Correcting personality tests for faking: A review of popular personality tests and an initial survey of researchers. *International Journal of Selection and Assessment*, 11, 340-344.
- Heggestad, E. D., & Gordon, H. L. (2008). An argument for context-specific personality assessments. *Industrial and Organizational Psychology*, 1, 320-322.
- Li, A., & Bagger, J. (2006). Using the BIDR to distinguish the effects of impression management and self-deception on the criterion validity of personality measures: A meta-analysis. *International Journal of Selection and Assessment*, 14, 131-141.
- McGill, D.A., van der Vleuten, C.P.M., & Clarke, M.J. (2011). Supervisor assessment of clinical and professional competence of medical trainees: A reliability study using workplace data and a focused analytical literature review. *Advances in Health Sciences Education*, 16, 405-425.
- McGrath, R.E., Mitchell, M., Kim, B.H., & Hough, L. (2010). Evidence for response bias as a source of error variance in applied assessment. *Psychological Bulletin*, 136, 450-470.
- Murphy, K.R. (2008). Explaining the weak relationship between job performance and ratings and job performance. *Industrial and Organizational Psychology*, 1, 148-160.
- Murphy, K.R., & Cleveland, J.N. (1995). *Understanding performance appraisal: Social, organizational, and goal-based perspectives*. Thousand Oaks, CA: Sage.
- Ones, D.S., Viswesvaran, C., & Reiss, A.D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology*, 81, 660-679.
- Paulhus, D.L. (1991). Measurement and control of response bias. In J.P. Robinson, P. Shaver, & L.S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17-59). San Diego, CA: Academic Press.
- Paulhus, D.L. (1998). *Manual for the Balanced Inventory of Desirable Responding (BIDR-7)*. Toronto, Canada: MHS.
- Paulhus, D.L. (2002). Socially desirable responding: The evolution of a construct. In H.I. Braun, D.N. Jackson, & D.E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 49-69). Mahwah, NJ: Erlbaum.
- Peck, J.A., & Levashina, J. (2017). Impression management and interview and job performance ratings: A meta-analysis of research design with tactics in mind. *Frontiers in Psychology*, 8:201.
- Shaffer, J.A., & Postlethwaite, B.E. (2012). A matter of context: A meta-analytic investigation of the relative validity of contextualized and noncontextualized personality measures. *Personnel Psychology*, 65, 445-494.
- Williams, K.M. (2022). The Occupational Performance Assessment – Response Distortion (OPerA-RD) scale: Measuring response bias in self- or other-reported job performance. *Journal of Personnel Psychology*, 21, 185-196.



RESEARCH ON CULTURALLY RELEVANT ITEMS IN EDUCATIONAL ASSESSMENTS

Deborah Bandalos

Deborah Bandalos

My APA colleagues are often surprised to learn that my background and training are not in psychology, but in education. Although I recently retired from the Department of Graduate Psychology at JMU, I spent much of my career in the Colleges of Education at the Universities of Nebraska and Georgia. Because of my background, I regularly attend the annual meetings of both APA and AERA, the American Educational Research Association, as well as that of the National Council on Measurement in Education (NCME). I bring all this up because NCME has recently hosted several webinars and conference sessions on the use of culturally relevant and justice-oriented test items in educational testing that I thought might be of interest to members of Division 5. I made that research the topic of my recent presidential address at the APA Convention and summarize it in this newsletter article.

Given the current political climate in which states such as Florida have implemented the so-called “Don’t Say Gay” bill and removed the College Board’s Advanced Placement course on African American History from the state’s curriculum, it may surprise some to learn that other school districts and K-12 testing organizations in the U.S. are experimenting with test items that focus on cultural relevance and social justice. Culturally relevant, or culturally responsive, test items have grown out of the movements for culturally relevant pedagogy (Ladson-Billings, 1994; 1995; 2014) and culturally responsive teaching (Gay, 2002; 2018). Ladson-Billings (1995) describes culturally relevant pedagogy as “...a theoretical model that not only addresses student achievement but also helps students to accept and affirm their cultural identity while developing critical perspectives that challenge inequities that schools (and other institutions) perpetuate” (Ladson-Billings, 1995, p. 469). Culturally responsive teaching incorporates the experiences, perspectives, and cultural milieu of ethnically diverse students into all aspects of teaching with the goal of creating a learning experience that is more personally relevant, meaningful, interesting, and thus more easily learned. It is thought that such a learning experience will, in turn, result in higher academic achievement.

The development of culturally relevant and justice-oriented test items is based on three tenets of culturally relevant pedagogy: student achievement, cultural competence, and sociopolitical consciousness. In culturally relevant pedagogy, measurement of student achievement should be designed to reveal each student’s growth, rather than comparing students to each other. Teachers are encouraged to use different types of assessments, reflecting the understanding that students learn in different ways. Finally, culturally relevant teaching is designed to infuse aspects of students’ cultural backgrounds into instruction through culture-based examples and references, and by asking questioning that require students to consider situations from the perspectives of those with different sociopolitical viewpoints. Teachers can also enhance students’ sociopolitical consciousness through assignments requiring critiques of social inequities.

Antiracist assessment practices Randall (2021; Randall et al. 2022) are based on the argument that traditional approaches to achievement testing perpetuate racism through their exclusive focus on white content and ways of knowing. As Randall notes (p. 5), “Issues of inequity and injustice are not to be resolved in assessment simply by increasing ‘access’ to whiteness (i.e., opportunity to learn) to marginalized students,” because such content ignores the culture and experiences of large segments of the student population. Randall advocates the development of test items that are explicitly designed to disrupt or expose racist ideas. Such test items might, for example, confront negative stereotypes about minorities or be based on contexts that highlight inequities. Examples of antiracist items are not yet widely available, but recent NCME webinars and conference presentations, as well as one recent doctoral dissertation have provided examples of culturally relevant and justice-oriented items and students’ reactions to these. Although I cannot do justice to this research in this short article, I hope the examples and student reactions shared here will generate interest. Steedle and Anguiano-Carrasco (2023) provided the example below of a culturally relevant test item developed at the ACT:

The lyrics of Lin-Manuel Miranda’s Hamilton soundtrack are sung or rapped at an average rate of 144 words per minute. Given that the Hamilton soundtrack is 2 hours 23 minutes long, which of the following is closest to the total number of words in the Hamilton soundtrack?

- A) 2,500
- B) 3,500
- C) 19,000
- D) 20,500
- E) 32,000

Patterson (2023) created the justice-oriented item below for his doctoral dissertation research on college students’ perceptions of such items in the context of an undergraduate assessment of information literacy:

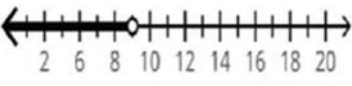
Shanice Richards wants to raise awareness of housing segregation. To inform her actions, she needs to research the history of housing segregation in the United States. Which of the following sources would be most useful for her to use?

- A) A website by the Shenandoah Valley Black Heritage Project that features housing segregation
- B) A website by a JMU professor that studies history
- C) A chapter from *The Color of Law*—a book written by scholars in race and housing
- D) A brochure from a realtor on houses in Harrisonburg

Randall (2021, p. 2) provides the example below of an antiracist math item:

Marcellus is cooking hot meals to hand out to a small group of twelve Black Lives Matters protesters demonstrating against separating families held at the U.S./Mexico border. He is making a meal of rice, cornbread, and red beans. He wants to make enough red beans for each person to have more than $\frac{3}{4}$ cup. Determine whether each inequality or number line correctly models c , the number of red beans Marcellus needs to make.

Select Yes or No for each option.

$c > 9$ ☒ Yes ☐ No
 $c < 9$ ☐ Yes ☒ No
 $c \geq 9$ ☐ Yes ☒ No
 ☐ Yes ☒ No
 ☐ Yes ☒ No
 ☒ Yes ☐ No

One question of interest to many researchers is how students react to these types of items. Do such items result in the hoped-for perceptions of more personal relevance, meaningfulness, interest, motivation, and engagement? If so, do these benefits result in higher academic achievement? Before attempting to answer these questions, I want to note that at a recent conference there was considerable pushback on my suggestion that more research was needed on these questions. Several audience members pointed out that we should not need research to support the inclusion in assessments of all members of society. Instead, the need for such inclusion should be taken as self-evident, thus negating the need for justification. I completely agree, but unfortunately, given today's political climate it seems inevitable that such questions will be asked.

To date, research on these questions has been qualitative, consisting mainly of cognitive interviews and focus groups with students of various ages. This research is focused on students' perceptions of the items rather than students' achievement on such tests. The quotations below are taken from the NCME conference and webinar presentations I alluded to previously. When available, I included information on student ethnicity and grade level. Overall, students appear to have positive reactions to the items, as evidenced in the quotations below from research conducted at WestEd and reported by Faulkner-Bond (NCME, 2022) as part of the NCME Equity Webinar series:

"I'm not sure if everybody had similar questions to look at, but I really did appreciate the diversity. I don't want to say 'I'...but most of the time people don't really get to see a whole bunch of [representation] and just thought it was really cool to include everybody instead of just having one specific group of people." (8th grade student)

Such support was voiced by both students of color and white students, but younger students seemed less aware of the items' cultural content. The comment below about a culturally relevant item was made by a white undergraduate student in Patterson's (2023) dissertation study:

"Okay. I liked it. I feel like I'm getting a sense of someone's life that's not mine. And it's interesting just seeing all of these different sources that all relate to the same topic but can go in different ways of how you perceive it and the different information that it gives about the item."

Although students did appreciate the diversity of the items, many students seemed to care more about the personal interest items held for them than their diversity per se. The quotation below is from a middle school student in one of Faulkner-Bond's (NCME, 2022) focus groups:

"I don't know if this is, like, the theme or not but I would not, like, change how interesting the questions are for diversity. I would rather read something more interesting than something I was included in. It's a lot more enjoyable for me to read something that's interesting and less lengthy."

The reference to question length in the quotation above was a common theme. Students tended to prefer shorter questions, even if this was at the expense of diversity. The quotation below is from Steedle and Anguiano-Carrasco's (2023) research on ACT items incorporating cultural relevant items:

"...if it's like a two sentence question or a four sentence cultural question I'd say go for the two sentence question... but if the questions are going to be the same length either way, I don't think the cultural content is what is slowing us down." (white 12th grade student)

Another student in Steedle and Anguiano-Carrasco's (2023) study addressed the motivational value of culturally relevant items:

"I think because...assuming what...the mindset of students who aren't too fond of taking the test is, I feel like they are just more like 'oh why is a richer you know Caucasian people making these tests for like me?' I feel like them knowing...people who are higher up care about their culture, will like motivate them to wanna do better for their own culture, like seeing their culture represented...can be very eye opening. It can be...life changing for a student to see...a question like that, and then for them to be like 'omg...they care about my culture...let me do this for my culture and for myself.'" (Latinx 11th grade student)

In his dissertation study, Patterson (2023) created both traditional and culturally relevant or justice-oriented versions of test items for an undergraduate test of information literacy. He then conducted focus groups with undergraduate students in which he asked about their perceptions of the items. One student commented that the items made them feel "seen:"

"[These culturally relevant items] made me feel represented in testing, which is something I didn't think that I wanted until I got it." (mixed race, transgender college student).

The qualitative research to date suggests that culturally relevant and justice-oriented items are generally perceived positively by students, with older students expressing more support for such items than younger students. White students were not always as enthusiastic about the new item types as were students of color and at least one student expressed difficulty in answering items with which they did not have direct experience, as evidenced by the comment below from a white student in Patterson's (2023) dissertation study:

"I think it's harder to answer these questions. I'm saying this because I've never experienced it. So, it's just harder to, I guess, answer them sometimes."

In Patterson's (2023) dissertation study, some students had trouble with diverse names for people and places. This comment illustrates the possibility that inclusion of such items could result in construct irrelevant variance due to unfamiliar terminology. However, as Randall, Slomp, Poe, and Oliveri (2022, p. 4) note that "In some respects, then, it can be said that what is construct-relevant behavior is dependent on the eye of the beholder.

What is construct irrelevant for the assessment designer, may very well be construct relevant for the examinee.” Construct relevance depends on the definition of the construct. If this definition includes students’ awareness of cultural factors or social justice issues, such material becomes construct relevant. A related issue is whether culturally relevant or justice-oriented contexts make items more cognitively complex. Such content can make items longer, as noted by Steedle and Anguiano-Carrasco (2023) and Patterson, among other researchers. Emotionally laden contexts might also be the source of construct irrelevant variance. Patterson investigated this issue and found that some students did express anger or other negative emotions on items related to unjust situations. For example, in the context of an item about the death or disappearance of indigenous women, one student stated:

“The emotion that I felt was anger because, again, it's not something that's talked about. It's because it's a group of minorities, and like, at least in this country, unless it's a white person it's not really looked into.” (Latina undergraduate student)

As a measurement professional who supports greater equity in assessment, I applaud the efforts of those developing culturally relevant and justice-oriented assessments and look forward to wider development and adoption of these. My hope is that, in the not-too-distant future, such assessments will be taken as a matter of course by both students and the public.

References

- Gay, G. (2002). Preparing for culturally responsive teaching. *Journal of Teacher Education*, 53(2), 106-116. <https://journals.sagepub.com/doi/pdf/10.1177/0022487102053002003>
- Gay, G. (2018). *Culturally responsive teaching: Theory, research, and practice* (3rd ed.). Teachers College Press.
- Ladson-Billings, G. (1994). *The dreamkeepers: Successful teaching for African-American students*. Jossey-Bass.
- Ladson-Billings, G. (1995). Toward a theory of culturally relevant pedagogy. *American Educational Research Journal*, 32(3), 465-491. <https://doi.org/10.3102/00028312032003465>
- Ladson-Billings, G. (2014). Culturally relevant pedagogy 2.0: A.k.a. the remix. *Harvard Educational Review*, 84(1), 74-135. <https://doi.org/10.17763/haer.84.1.p2rj131485484751>
- National Council on Measurement in Education. (2022, January 19). *Cultural relevance versus construct relevance: How do we create culturally responsive assessments?* [YouTube].
- Patterson, C. R. (2023). Observing how college students process culturally responsive test items. [Unpublished doctoral dissertation]. James Madison University.
- Randall, J. (2021). Color-neutral is not a thing: Redefining construct definition and representation through a justice-oriented critical antiracist lens. *Educational Measurement: Issues and Practice*, 40(4), 82-90.
- Randall, J., Slomp, D., Poe, M., & Oliveri, M.E. (2022). Disrupting white supremacy in assessment: Toward a justice-oriented, antiracist validity framework. *Educational Assessment*, 27(2), 170-178. <https://doi.org/10.1080/10627197.2022.2042682>
- Steedle, J., & Anguiano-Carrasco, C. (2023, March 30). Developing culturally relevant math and science items: Lessons learned and student reactions. [Conference presentation]. Annual Meeting of the National Council on Measurement in Education, Chicago, IL.



DIVISION 5 AWARDS 2023

*James Christopher Head**James Christopher Head*

Division 5 of APA offers eight annual awards to recognize the contribution of members to qualitative and quantitative methods in psychology. Nominations for awards are posted in The Score, the SQIP newsletter, and the Division 5 website. This year, five awards were conferred. Awardees were recognized at the awards ceremony organized by Division 5 and held in conjunction with the APA 2023 Annual Convention.

Qualitative Awards

Committee: James Christopher Head (chair), Linda McMullen, and Deborah Tolman

Distinguished Dissertation in Qualitative Inquiry Award Winner:



Edward D. Scott, Jr., PhD

Dr. Scott's dissertation, *Making Justice: A multimethod qualitative study of activist identity formation and political imagination(s) among Black adolescents*, is clear, cohesive, elegant, and compelling. After articulating a past-present-possible conceptual framework, Scott used narrative and arts-based methods to interrogate the intersections of race, youth development, identity, and activism as he examined the sociopolitical development *activist identity* in adolescents. In the words of Nancy L. Deutsch, Scott employed "rigorous and innovative qualitative methods, in both design and application, resulting in a truly superb dissertation that is a delight to read, is intellectually rigorous, and is an exemplar of qualitative methods."

Distinguished Early Career Contributions in Qualitative Inquiry Award Winner:



Roberto Luis Abreu, Ph.D.

Since earning his Ph.D. in 2018, Roberto Luis Abreu has made valuable contributions to qualitative inquiry through his scholarship and teaching. He has published extensively, has received many awards for his scholarship, and promotes qualitative inquiry in a variety of scholarly communities. In the words of Dr. Heidi Levitt, Dr. Abreu “is a committed scientist-practitioner-advocate who has advanced qualitative methods and already is invested in contributing towards future generations of qualitative and mixed methods researchers.”

Quantitative Awards

Committee: Qiwei He (chair), Emily Shaw and Rochelle Michel, with support from David Budescu and Amy Schmidt

Anne Anastasi Dissertation Award Winner:



Kenneth Tyler Wilcox, Ph.D.

Dr. Kenneth Tyler Wilcox received his Ph.D. in 2022 from University of Notre Dame in Quantitative Psychology. His dissertation, *Jointly modeling participant-level data and summary statistics for treatment*, is novel and important. His developed Bayesian modeling approaches for cumulative data analysis using both participant-level data and aggregated data significantly adds to the literature of cumulative data analysis methodology. Dr. Wilcox’s dissertation is of high practical implication to enable psychologists to better utilize all available quantitative information for cumulative data analysis. His dissertation well presents far-reaching method in terms of utility across disciplines and has been published in top-tier journals such as *Psychological Methods* and *Multivariate Behavioral Research*. In addition, Dr. Wilcox also extends his dissertation work to two promising new research lines in topic modeling and text mining.

Anne Anastasi Distinguished Early Career Contribution Award:



Heining Cham, Ph.D.

Dr. Cham received his Ph.D. in 2013 in Psychology in Quantitative Research Methods from Arizona State University and is currently Associate Professor (with tenure), of Psychometrics and Quantitative Psychology, in the Department of Psychology at Fordham University. Since completing his doctoral studies, he has made numerous important contributions to the field of quantitative psychology and beyond. Dr. Cham’s quantitative research contributions focus on quasi-experimental design and causal inference and latent variable interactions. Running through his research in both areas is a focus on the critical issue of the impact of missing data on the behavior of statistical estimation procedures. Dr. Cham’s reach in quantitative methods goes well beyond his core quantitative methods focus. He is also a prolific collaborator, and with substantive colleagues, has made

important contributions to research including advanced research design, statistics and assessment methodology in service of answering critical questions concerning health, mental health, and human development.

Jacob Cohen Award for Distinguished Contributions to Teaching and Mentoring:



Viji Sathy, Ph.D.

Dr. Viji Sathy is Associate Dean for Evaluation and Assessment, Office of Undergraduate Education, College of Arts & Sciences, and Professor of the Practice of Psychology and Neuroscience in University of North Carolina at Chapel Hill. Dr. Sathy has made unique and outstanding contributions to the teaching and mentoring of quantitative psychology through her scholarship and practice. She has provided high quality teaching in hundreds of statistics and research methods courses; created high structure, active learning environments; mentored large graduate student teams of teaching assistants; and successfully engaged colleagues in teaching practice to celebrates classroom diversity and ensures that students with different backgrounds are included in the learning process. Her contributions are both micro and macro with a broad impact not only to her over thousands of students but also the entire field. Over her career starting in quantitative psychology and broadening to the scholarship of inclusive teaching and learning, Dr. Sathy made numerous clear and outstanding professional contributions that broadly affect how professors teach (including statistics and research methods) and our understanding of how diverse students are supported in our measurement and statistics courses, and has been successfully, thoughtfully, and innovatively doing so. Dr. Sathy brings the best practices and evidence from educational research in pedagogy to her classroom and guides students through the learning process with her best care.

Note: No nominations were received for the Samuel J. Messick Award for Distinguished Scientific Contributions, the Distinguished Contributions in Qualitative Inquiry Award, or the Distinguished Contributions to Teaching and Mentoring in Qualitative Inquiry Award.

Next Year's Awards:

The deadline for next years' awards will be the first Monday in December (December 4, 2023). I encourage you to mark your calendars, spread the word, and consider nominating a recent graduate, early career professional, or other well-deserving colleague.



WHAT'S NEW IN BOOKS

Compiled by
Joanna Solomon

“An Introduction to Artificial Psychology: Application Fuzzy Set Theory and Deep Machine Learning in Psychological Research using R” (<https://link.springer.com/book/10.1007/978-3-031-31172-7>)

By Hojjatollah Farahani, Marija Blagojević, Parviz Azadfallah, Peter Watson, Forough Esrafilian, and Sara Saljoughi

Published May 2023 by Springer (\$199.00 hardback; e-books also available)

This text combines fuzzy set theory with psychology research to introduce a modern, robust analysis method. The real world is complex, and traditional statistics are often limited, so this novel methodology can offer additional utility and applications for multidisciplinary research. Quantitative variable-oriented methodology and qualitative case-oriented methods are both used to explain the set-oriented methodology and this book combines the precision of quantitative methods with information from qualitative methods. Chapter topics include cognitive maps, methodology, network analysis, neural networks, Bayesian methods, and feature selection. Examples, data, and sample code are provided for R software.

“Qualitative Data Analysis: Key Approaches” (<https://us.sagepub.com/en-us/nam/qualitative-data-analysis/book271012>)

By Peter A. Stevens

Published January 2023 by SAGE (\$142.00 hardback; e-books also available)

With a clear focus on the practical, how-to angle, this book provides researchers, analysts, and students with a guide for conducting and interpreting qualitative data analysis. The book is designed to be accessible, thorough, and applicable to a wide array of disciplines within the social sciences. Contents include text analysis, narrative analysis, investigating causal relationships, and thematic analysis. The book is designed as a companion for users of NVivo software, a commonly used software for qualitative work, which will support new and experienced researchers alike.

“An Introduction to Psychometrics and Psychological Assessment: Using, Interpreting and Developing Tests (2nd edition)” (<https://www.routledge.com/An-Introduction-to-Psychometrics-and-Psychological-Assessment-Using-Interpreting/Cooper/p/book/9781032146171>)

By Colin Cooper

Published May 2023 by Routledge (\$128.00 hardback; \$42.36 paperback; e-books also available)

This book offers an expansion and update to a previous book, *Psychological Testing: Theory and Practice*. The author explores how questionnaires and tests are developed, administered, scored, and interpreted. Intended to be appropriate for test users, test developers, assessment researchers, and students, it provides readers with a thorough discussion of the limitations and guiding principles of tests, theory, and assessment methodology. It provides an evaluation of different approaches, discussing in detail the strengths and drawbacks of various methods. It is an introductory text, but it touches on a wide variety of topics, including cutting-edge techniques and best practices.

“Machine Learning for Social and Behavioral Research” (<https://www.guilford.com/books/Machine-Learning-for-Social-and-Behavioral-Research/Jacobucci-Grimm-Zhang/9781462552924>)

By Ross Jacobucci, Kevin J. Grimm, and Zhiyong Zhang

Published July 2023 by Guilford Press (\$93.00 hardback; \$62.00 paperback; e-books also available)

In this latest book in the *Methodology in the Social Sciences* series by Todd D. Little, readers will learn skills needed to analyze large, complex data, report and interpret results, and make use of machine learning tools. The techniques discussed rely on practical examples from a variety of disciplines within psychology to teach readers theory and applied use. Topics covered include measurement error, regularization, text and network data, decision trees, alongside methodological approaches concerning univariate and multivariate data. Chapters include sections regarding discussions of key R packages; the companion website provides R programming scripts and data for the book's examples.

Guidelines for the “What’s New?” Column:

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 the What’s New 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 *The Score* Associate Editor, Joanna Solomon: joanna.solomon@mhs.com.



PUBLICATON CORNER

Compiled by
Gill Sitarenios

A Review of Applications of the Bayes Factor in Psychological Research

This article reviews applications of the Bayes Factor over the last 25 years. Bayesian approaches in research have shown pronounced attention for the Bayes factor as a tool for hypothesis evaluation and model selection. Six main applications are described including Bayesian evaluation of point null, interval, and informative hypotheses, Bayesian evidence synthesis, Bayesian variable selection and model averaging, and Bayesian evaluation of cognitive models. Illustrative examples are provided, as well as an overview of key references and software with links to other applications. The article concludes with a discussion of the opportunities and pitfalls of Bayes factor applications and future lines investigation.

Heck, D. W., Boehm, U., Böing-Messing, F., Bürkner, P.-C., Derks, K., Dienes, Z., Fu, Q., Gu, X., Karimova, D., Kiers, H. A. L., Klugkist, I., Kuiper, R. M., Lee, M. D., Leenders, R., Leplaa, H. J., Linde, M., Ly, A., Meijerink-Bosman, M., Moerbeek, M., . . . Hoijtink, H. (2023). A review of applications of the Bayes factor in psychological research. *Psychological Methods*, 28(3), 558-579. <https://doi.org/10.1037/met0000454>

An Introductory Guide for Conducting Psychological Research with Big Data

This article provides an introductory guide for conducting Big Data research for psychologists who are considering using this approach. The authors take the Knowledge Discovery from Database steps as illustrative, and show how it provides useful indications for finding data suitable for psychological investigations. They describe how these data can be preprocessed, and list some techniques to analyze them and through programming languages like R and Python. They also explain the concepts with the terminology and take examples from psychology. This overview helps build a general insight into the research steps and a common language, facilitating collaboration across different fields.

Vezzoli, M., & Zogmaister, C. (2023). An introductory guide for conducting psychological research with big data. *Psychological Methods*, 28(3), 580-599, <https://doi.org/10.1037/met0000513>

Summary-Statistics-Based Power Analysis: A New and Practical Method to Determine Sample Size for Mixed-Effects Modeling

This article proposes a summary-statistics-based power analysis for conducting power analysis for mixed-effects modeling with two-level nested data. This approach complements the existing formula-based and simulation-based methods. The proposed method is based on the conditional equivalence of the summary-statistics approach and mixed-effects modeling, paring back the power analysis for mixed-effects modeling to that for a simpler statistical analysis (e.g., one-sample t test). Analytic proof

is provided through a series of statistical simulations to show the validity and robustness of the summary-statistics-based power analysis and show illustrative examples with real published work. A web app was also developed to facilitate the utility of the proposed method. This approach provides an alternative for applied researchers when there is limited information to conduct power analysis.

Murayama, K., Usami, S., & Sakaki, M. (2022). Summary-statistics-based power analysis: A new and practical method to determine sample size for mixed-effects modeling. *Psychological Methods*, 27(6), 1014-1038. <https://doi.org/10.1037/met0000330>

Sample Size Calculation and Optimal Design for Regression-Based Norming of Tests and Questionnaires

Greater precision in test norms can be achieved by drawing a large normative sample and using regression-based norming. Based on that norming method, this article presents a procedure for sample size planning to make inference on Z-scores and percentile rank scores. Sampling variance formulas for these norm statistics are derived and used to obtain the optimal design, that is, the optimal predictor distribution, for the normative sample, thereby maximizing precision of estimation. This approach is applied to five regression models with a quantitative and a categorical predictor, differing in whether they allow for interaction and nonlinearity. Formulas are provided to compute the normative sample size such that individuals' positions relative to the derived norms can be assessed with prespecified power and precision.

Innocenti, F., Tan, F. E. S., Candel, M. J. J. M., & van Breukelen, G. J. P. (2023). Sample size calculation and optimal design for regression-based norming of tests and questionnaires. *Psychological Methods*, 28(1), 89-106. <https://doi.org/10.1037/met0000394>

A Tutorial on Assessing Statistical Power and Determining Sample Size for Structural Equation Models

This manuscript provides a step-by-step illustration of how *a priori*, *post hoc*, and *compromise power analyses* can be conducted for a range of different SEM applications. Using illustrative examples the authors demonstrate power analyses for hypotheses regarding overall model fit, global model comparisons, particular individual model parameters, and differences in multigroup contexts.

Jobst, L. J., Bader, M., & Moshagen, M. (2023). A tutorial on assessing statistical power and determining sample size for structural equation models. *Psychological Methods*, 28(1), 207-221. <https://doi.org/10.1037/met0000423>

Guidelines for the “Publication Corner” Column:

Have you seen a recent publication that you think would be of great interest to Division 5 membership? Please feel free to share suggestions for publications to include in this column. Please send your ideas to *The Score* Editor, Gill Sitarenios (Gill.Sitarenios@gmail.com). Please be sure to include full reference details and a short synopsis (2-3 sentences) of the publication. Use “Publication Corner” in the subject line of the email.

LINKS TO RECENT PREVIOUS ISSUES OF “THE SCORE”



Issue Date	Vol #	Issue #	Link
July, 2023	45	3	
April, 2023	45	2	https://www.apadivisions.org/division-5/publications/score/2023/04/issue.pdf
January, 2023	45	1	https://www.apadivisions.org/division-5/publications/score/2023/01/issue.pdf
October, 2022	44	4	https://www.apadivisions.org/division-5/publications/score/2022/10
July, 2022	44	3	https://www.apadivisions.org/division-5/publications/score/2022/07
April, 2022	44	2	https://www.apadivisions.org/division-5/publications/score/2022/04
January, 2022	44	1	https://www.apadivisions.org/division-5/publications/score/2022/01
October, 2021	43	4	https://www.apadivisions.org/division-5/publications/score/2021/10
July, 2021	43	3	https://www.apadivisions.org/division-5/publications/score/2021/07
April, 2021	43	2	https://www.apadivisions.org/division-5/publications/score/2021/04
January, 2021	43	1	https://www.apadivisions.org/division-5/publications/score/2021/01
October, 2020	42	4	https://www.apadivisions.org/division-5/publications/score/2020/10
July, 2020	42	3	https://www.apadivisions.org/division-5/publications/score/2020/07
April, 2020	42	2	https://www.apadivisions.org/division-5/publications/score/2020/04
January, 2020	42	1	https://www.apadivisions.org/division-5/publications/score/2020/01

Issues of The Score can be accessed at the following webpage:

<https://www.apadivisions.org/division-5/publications/score/issues>

DIVISION 5 MEMBERSHIP INFORMATION



Division 5 draws together specialists in the areas of quantitative and qualitative methods. Membership includes specialists in the fields of educational psychology, industrial-organizational psychology, individual assessment, measurement, program evaluation, psychometrics, qualitative inquiry, research interviewing, research methods and statistics. Division 5 provides the opportunity to participate in this critical aspect of professional development, and is a vibrant community that serves the core interests and needs of rigorous scientific inquiry in the social sciences. Those who are interested in joining the division or who would like more information should be referred to the Division 5 website:

<https://www.apadivisions.org/division-5/membership>

Advertising in *The Score*

The Score is the newsletter of the American Psychological Association's Division 5—Evaluation, Measurement, and Statistics. 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 Gill Sitarenios at GSitarenios@yahoo.ca.

Size	Dimensions	Display Ad Price	Job Announcement Price
Full page	7.125" × 9.5"	\$235	Not available
Half page	7.125" × 4.75"	\$175	\$120
Third page	4.75" × 4.75"	\$125	\$90
	2.375" × 9.5"		
Sixth page	2.375" × 4.75"	\$90	Free/\$55
Notes: Insertion orders for four consecutive issues receive a 15% discount. First sixth page job ad each year free, thereafter \$55.			