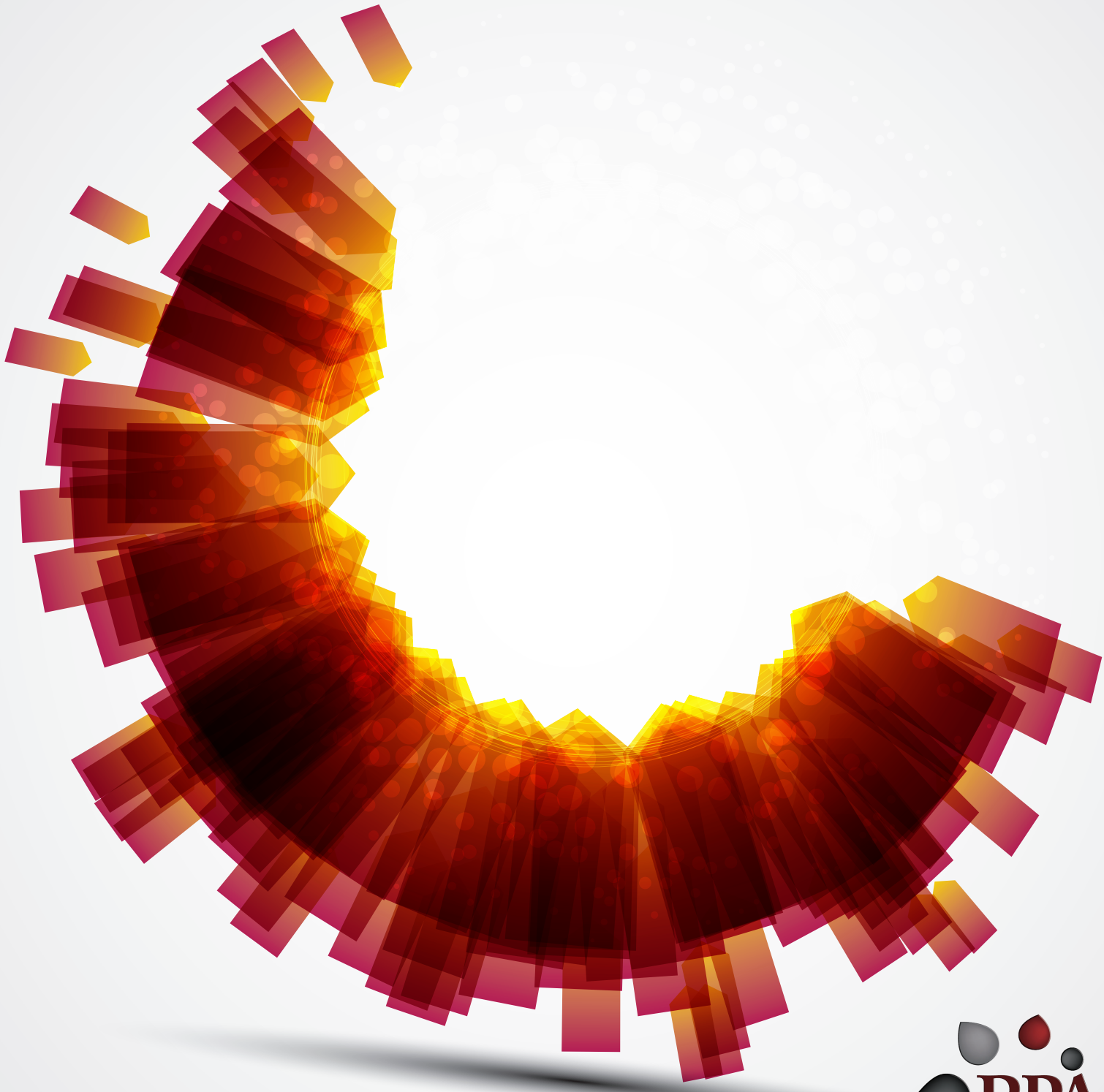
RESEARCH & PRACTICE IN ASSESSMENT

VOLUME SEVEN | SUMMER 2012
www.RPAjournal.com
ISSN # 2161-4210



RPA

A PUBLICATION OF THE VIRGINIA ASSESSMENT GROUP

RESEARCH & PRACTICE IN ASSESSMENT

The goal of *Research & Practice in Assessment* is to serve the assessment community as an online journal focusing on higher education assessment. It is dedicated to the advancement of scholarly discussion amongst researchers and practitioners in this evolving field. The journal originated from the Board of the Virginia Assessment Group, one of the oldest continuing professional higher education assessment organizations in the United States. *Research & Practice in Assessment* is a peer-reviewed publication that uses a double-blind review process. Approximately fifty percent of submissions are accepted for issues that are published twice annually.

TABLE OF CONTENTS

CALL FOR PAPERS

Research & Practice in Assessment is currently soliciting articles and reviews for its Winter 2012 issue. Manuscripts submitted to RPA may be related to various higher education assessment themes, and should adopt either an assessment measurement or an assessment policy/foundations framework. Contributions to be evaluated by the RPA Review Board should be submitted no later than September 1. Manuscripts must comply with the RPA Submission Guidelines and be sent electronically to:

rpaeditor@virginiaassessment.org

Our journal is available for free download for iPad!
www.RPAjournal.com

Published by:
VIRGINIA ASSESSMENT GROUP
www.virginiaassessment.org

For questions about Virginia Assessment Group membership opportunities email webmaster@virginiaassessment.org. We welcome members and non-members to join our community. If you would like to be informed of Virginia Assessment Group events, please contact us and we will add you to our distribution list.

Publication Design by Patrice Brown
www.treestardesign.com

FROM THE EDITOR

Conceptually, this issue of *Research & Practice in Assessment* begins with its end. The concluding Ruminate section highlights an inter-cultural fable, "The Blind Men and the Elephant." Nineteenth century poet, John Godfrey Saxe, penned the rendition that is familiar to most Western cultures:

> *It was six men of Indostan to learning much inclined,*
> *Who went to see the Elephant (though all of them were blind),*
> *That each by observation might satisfy his mind.*

In the stanzas that follow, Saxe offers readers the observations and corresponding assessments made by each blind man. Two men assert that the matter is "mighty plain," one "bawls" his deduction aloud, and three others conclude they are able to "see." One subject. Six individuals. Six perspectives. Similarly, the current issue puts forth intentionally diverse views on higher education assessment. Recent structural changes to the Editorial and Review Boards are designed to sustain diversity of thought, and promote rich assessment discourse among colleagues.

This issue opens with a provocative special feature penned by Wake Forest University sociologist Joseph Soares, who argues for the further development and use of predictively powerful tests that lack social prejudice. The piece is adapted from his latest book, *SAT Wars*, an edited volume that examines social effects of high-stakes standardized testing. Three peer-review articles follow. Zilberberg, Anderson, Swerdzewski, Finney, and Marsh address the impact of college students' understanding of federal accountability testing and their corresponding testing behaviors. From there, Hoffman and Bresciani examine knowledge, skill, and dispositional competencies among student affairs professionals. Finally, Erwin employs a longitudinal design to link alumni self-ratings of personal growth with intellectual development.

In the latter half of the issue, I encourage readers to peruse the Review and Notes In Brief sections. Lagotte offers commentary on *Good Education in an Age of Measurement*, a penetrating work for assessment professionals. Mahiri's review of *Science Learning and Instruction* revisits the topic of knowledge integration and, in the first *RPA* software review, Gotzmann and Bahry focus on the free item analysis application *jMetrik*. Within the Notes In Brief section, practitioners and scholars alike may appreciate Zelna and Dunstan's annotated list of selected assessment conferences. Ruminate closes this diverse issue (as it began) with a symbiotic display of image and prose by Basbagill and Saxe.

In the past six months, *RPA* has taken yet another qualitative leap through publication improvements and website development. I am indebted to various leaders of the Virginia Assessment Group - past and present- who provided the necessary resources and assistance to accomplish these changes: Kathryne Drezek McConnell, Keston Fulcher, Robin Anderson, and the current board members. The comprehensive website redesign could not have been accomplished without the *exceptional* talents of Katelynn Stein and Patrice Brown. A final note of commendation is in order for the *RPA* Editorial and Review Board members; your level of involvement and mentoring during the peer review process has been admirable.

As you engage the pieces contained herein, I hope you will consider penning your own scholarly piece for submission to *Research & Practice in Assessment.*

Regards,

Joshua T. Brown

Liberty University

Joseph Soares is Professor of Sociology at Wake Forest University. This article is adapted from his latest book, *SAT Wars: The Case for Test-Optional Admissions* (Teacher's College Press, 2011) an edited volume that examines the social effects of high-stakes standardized testing. Additional contributors include Richard Atkinson, Thomas Espenshade, Daniel Golden, Charles Murray, and Robert Sternberg, among others.

AUTHOR

Joseph A. Soares, Ph.D.
Wake Forest University

CORRESPONDENCE

Email
soaresja@wfu.edu

For Tests that are Predictively Powerful and Without Social Prejudice

In Philip Pullman's dark matter sci-fi trilogy[1], there is a golden compass that in the hands of the right person is predictively powerful; the same was supposed to be true of the SAT/ACT – the statistically indistinguishable standardized tests for college admissions. They were intended to be reliable mechanisms for identifying future trajectories, not unlike a meritocratic fortune telling device. In Pullman's novels, the compass works; however in the real world, the predictive accomplishments of the SAT/ACT are sadly less dramatic.

Pullman's novels also posit the existence of multiple parallel universes where enlightenment and love struggle against dogma and hate. If multiple universes exist, surely some of them by now have worked out how to make college admissions meritocratic, for even we are approaching that goal in the early part of our twenty-first century. We have labored long, misdirected by an old-compass admissions system, designed in the heyday of eugenics, which worked more effectively to exclude social "undesirables" than to include those who were academically fit. In the last two decades, however, nearly a third of our four-year-degree-granting institutions have gone "test-optional" breaking in part or whole with the old-compass camp. New tools, often called non-cognitive tests, which statistically outperform previous tests and do so without transmitting social disparities, have been used by thousands of students at universities as diverse as Tufts, DePaul, and Oklahoma State. Today, there are good reasons to be optimistic about the progress being made in the real world.

Test Scores Add Little to High School GPA

What I am referring to here as the old-compass admissions system "is the 20th century formula of looking at high-school record and one of two standardized tests, either the SAT or ACT, in order to predict grades in the first year of college" (Soares, 2012b, p.

66). The scientific prowess of the old method was never found to be very great, predicting at best, according to the test makers, about 21% of the variance in college grades (Kobrin, Patterson, Shaw, Mattern, & Barbuti, 2008). The contribution of each part of the old system, however, high school grades and test scores, was far from equal.

Though many parents and academics are surprised by this, it remains true that high school grades have always done a better job in predicting college grades than test scores. As the Emeritus President of the University of California, Richard Atkinson, and Berkeley statistician Saul Geiser, remind us, "Irrespective of the quality or type of school attended, cumulative grade point average (GPA) in academic subjects in high school has proved to be the best overall predictor of student performance in college. This finding has been confirmed in the great majority of 'predictive-validity' studies conducted over the years, including studies conducted by the testing agencies themselves" (Atkinson & Geiser, 2012, p. 24). In technical articles, for statistical *cognoscenti,* the College Board concedes that high school grades matter most, but for *hoi polloi* of the press, they go "truth optional" and unabashedly claim that the test predicts best (Kobrin et al., 2008; Morgan, 1989; for press coverage, see: http://thechoice.blogs.nytimes.com/2011/11/09/sat/).

When 70 to 80 percent of the variance in college grades is left unexplained by our best statistical models, it is time again to acknowledge that admissions professionals do not have a golden compass; they are making decisions that remain more art than science.

Because the SAT and ACT tests are less predictive than the high-school record, the real question is, how much value do they add? Youths and their families should not have to suffer through the time, expense, and effort to take a test that stands outside the high school curriculum, unless it raises to a higher level our ability to identify college-ready talent. When statisticians attempt to model outcomes such as SAT scores, which range from 200 to 2800, they use linear multiple regressions to measure the contribution that each variable makes to the explanatory power, or R-square, of the model. The test industry claims to find an 8-point boost, raising one's R-square from 13% with high school GPA alone, to 21% with the SAT (Kobrin et al., 2008). Independent researchers, however, most often find an increase of merely 2 points (Soares, 2012a). As one can see from examples in my book, *SAT Wars,* institutional validity studies show that the SAT increased Johns Hopkins' R-square by two percentage points, raising their models' explanatory punch from an R-square of 0.18 to 0.20; at the University of Georgia it added one percent, raising their R-square from 0.30 to 0.31; and at DePaul the ACT was found to contribute one percent, raising their R-square from 0.19 to 0.20. Independent scholars found that neither the SAT nor the ACT adds more than a few percentage points to what is already known from high school GPA. For a billion-dollar industry, this is pretty pathetic value added for the money.

If the SAT/ACT improves one's predictive model by just one or two percentage points, how could that be worth the costs? Those tests do not lift college admissions out of the realm of practical wisdom into the realm of applied science. When 70 to 80 percent of the variance in college grades is left unexplained by our best statistical models, it is time again to acknowledge that admissions professionals do not have a golden compass; they are making decisions that remain more art than science. A false sense of scientific precision is one type of collateral damage done by the test industry. When test scores are used to set floors below which admissions staff will not go, we are doing an injustice to thousands of students; and when we decide between students based on a test score difference, we are relying on a compass that cannot find true north.

Some Tests Calcify Social Disparities

In addition to being largely redundant with information provided by the high school transcript, these particular tests are discriminatory. Not all tests disguise social selection as academic selectivity, but the SAT and ACT do. Admissions by the old-compass method "narrows the socioeconomic and racial diversity of one's pool and yield. The more one relies on SAT/ACT/LSAT-type standardized tests, the more social disparities unfavorable to racial minorities, women, and low SES youths are passed along" (Soares, 2012b, p. 66). Those tests tell us that women are less quantitative than men, because females score on average 33 points lower than males on math sections. Hispanics/Mexican Americans and Blacks are "dumb and dumber," with the former falling 219 points, and the latter 303 points, on average behind Whites.

Test score disparities by gender and race do not end the list of demographic problems with the test. Family income has a strong linear relation to test score: the higher one's family's income, the higher the average test score. In fact, test scores correlate more strongly with family income than with high school grades. Students from poor families, those earning less than 20 thousand dollars annually, score 100 points lower than students from families earning near the median range in America, between 40 and 50 thousand dollars per year; and further, those students from median income families score 200 points behind students from families earning over 100 thousand dollars annually.

Some researchers have expressed the concern that HSGPA might be more correlated with family socioeconomic status (SES) than is the case for SAT scores (Sternberg, Bonney, Gabora, & Merrifield, 2012). An argument used to defend the SAT/ACT is that these tests level the playing field, providing for a nationally-normed test that reduces disparities among high schools due to the property values of the neighborhood and the SES composition of the student body. But University of California researchers found the opposite. Geiser and Santelices (2007) "reported that the SAT-V correlated at the .32 level with family income, and at the .39 level with parents' education; similarly, SAT-M scores correlated respectively at .24 and .32, but HSGPA correlated with family income at the .04 level, and with parents' education at the .06 level" (p. 2). If Geiser and Santelices are right, HSGPA is far from being a proxy for social class. Since HSGPA retains its punch, without conveying social disparities, then why not save money, energy, and incalculable family anxiety by dropping the SAT/ACT? (Soares, 2012b).

One indirect effect of the SES selection accomplished by using these tests is an economic payoff for institutions in higher education. Colleges can balance their budget with full-fare paying families if they can advertise high average test scores for admitted students.

One indirect effect of the SES selection accomplished by using these tests is an economic payoff for institutions in higher education. Colleges can balance their budget with full-fare paying families if they can advertise high average test scores for admitted students. The higher the college's average score, the more economically affluent the next year's applicant pool. Prospective students will self-select away from or toward institutions based on test scores, and in doing so assure that very selective colleges are economically homogenous and privileged. Bank accounts, not brains, determine which birds flock together. Needs-blind admissions furthers the SES charade, because only the most economically exclusive colleges can afford to bank on an applicant pool so affluent that it never risks admitting more needy students than it can afford to cover (Soares, 2007).
Some will say, if these tests select for youths from families with higher incomes, and against women, Hispanics, and Blacks, is that just a reflection of our society's inequalities in academic preparation? Is it not the case that White males from affluent families are going to receive the most resources and attention from their families and schools? Perhaps, the test is fair and the group disparities it displays are just a measure of life's unfairness?

I have already offered for your consideration Geiser and Santelices' (2007) finding that family income and parents' education correlate with test scores but do not correlate with grades earned in high school. From their work, one can see that SAT selection promotes social disparities not captured by selection based mainly on HSGPA. Selection by test scores stratifies higher education into a class system: the higher one's college's selectivity, the higher the SES composition of one's student body (Soares, 2007). Evidence is also available from the University of Texas, where the natural experiment of admitting all students in the top ten percent of each high school class were admitted, which enabled racial and social class diversity, without detriment to the students or the university.

Table 1
Percent of Each Higher Education Tier Occupied by Each SES Quartile

SES Quartiles	% of Tier 1	% of Tier 2	% of Tier 3	% of Tier 4	% of Tier 5	% of Tier 6	% of Tier 7
Top	79	64	51	37	23	36	22
Upper Middle	16	19	24	27	28	21	25
Lower Middle	3	9	14	23	28	24	27
Bottom	2	7	10	13	20	19	25

Source: National Educational Longitudinal Survey, 1988-2000. U.S. Department of Education. Restricted Access Data License Control Number: 06011044 (as cited in Soares, 2007).

As described in my book, *SAT Wars,* "The Vice-Provost [at the University of Texas in Austin] for admissions, Dr. Bruce Walker, has released multiple reports on the 10% solution showing how high school ranking is an excellent and reliable predictor of college GPA and graduation" (Soares, 2012a, p. 203). Class rank reduces, rather than passes along, SES disparities. "Being in the top 10 percent of any high school graduating class, allows a youth to overcome the disadvantages of coming from a low income family; of having parents without high school degrees; and of attending a low performing high school. Top 10 percent youths from families with the lowest incomes, below $20,000 per year, and from the least desirable high schools, those officially ranked "low performing," do better academically at the University of Texas than youths below the top 10 percent from "exemplary" high schools, who are from high-income families, and with college-educated parents" (p. 203).

If school grades and class rank are less influenced by SES than the SAT/ACT, the absence of SES effects on high school grades could be due to the stratification that divides students by race and class into different high schools in the first place. Again, one may still argue that the test is just a reflection of life's inequalities. But there is another, more sinister possibility. What if the test has social discrimination built into it? What if the questions used on the test systematically favor some groups over others?

Test Question Selection and Social Bias

What if the test has social discrimination built into it? What if the questions used on the test systematically favor some group over others?

In *SAT Wars*, Jay Rosner, the vice-president of the Princeton Foundation, offers shocking evidence of systematic bias in the SAT's logic of question selection. The questions that count on each year's version of the SAT are drawn from experimental questions that are pretested in previous years. Each test combines questions that will count for that year's scores and experimental questions that are being vetted to see how they perform for future use. The difference between a good experimental question and a bad one is whether it retains the bell curve shape of test score results. The SAT has retained the same bell curve distribution ever since 1926, which some take as a measure of its validity, rather than as an indicator of its role in transmitting social disparities. Working with two years of national SAT data, Rosner (2012) found there are few "neutral" test questions, in the sense that both men and women, Blacks and Whites, all perform equally well or equally poorly on them. Rather, all but one or two questions in each section of the real test are questions that, when they were rolled out in the experimental section of previous tests, students performed differently on those experimental questions based on their demographic profile: race, gender, and family income. Students taking the test are invited, for reasons of research, to voluntarily provide demographic information on themselves. Rosner presents in chapter six of SAT Wars examples of math questions that women, and verbal questions that Blacks and Hispanics, outperform males and Whites on, respectively.

Here is one example of a verbal sentence completion question that produces a racial test score gap: "The actors bearing on the stage seemed _____; her movements were natural and her technique _____." Rosner then provides the five possible word-combination answers that were used on the SAT, tells the reader that the correct answer was "(C) unstudied … uncontrived" and invites the reader to guess whether this was a question Whites outperformed Blacks on or the reverse. One may think this looks like a good question, using terms that belong in a college student's vocabulary, but that is incorrect. Rosner informs us that this is a Black advantage question, on which Black youths outperform Whites; and because of that, it does not make it onto next year's SAT exam. This question never counted. Rosner finds that out of the 156 verbal questions on two years of the SAT that counted, zero were questions like the one above, on which Blacks scored better than Whites (Rosner, 2012). All verbal questions on the SAT have been White advantage questions. I am not going to provide additional examples of racial bias or gender bias in the question selection step for the SAT, because I would like you all to read Rosner's contribution. But I will say that if I were able to pick next year's questions, rather than rely on a statistical algorithm that retains a bell curve, I could eliminate the test's gender gaps on math scores and racial gaps on verbal scores.

Non-Cognitive Tests Predict Better, and Without Social Prejudice

To those who still believe that test score disparities by demographic groups are just a reflection of life's unfairness, I would point to the chapter in *SAT Wars* written by Robert Sternberg. Sternberg was dean at Tufts University when that institution adopted the Kaleidoscope project to assess applicants' creative and practical problem solving ability. They found these "non-cognitive" tests performed statistically better than the SAT in predicting grades and college retention; and they did so without any gender or racial test score gaps. The January-March 2012 issue of *Educational Psychologist* provides case studies of non-cognitive tests for undergraduate admissions and for law school admissions that are simultaneously predictively more powerful and without transmitting the social disparities of the SAT or LSAT. There are tests that predict without prejudice. We are not inescapably compelled to transmit society's previous social inequalities.

I find myself arguing that the SAT should be ended. Not deemphasized, but no longer administered… [T]he SAT score, intended as a signal flare for those at the bottom, has become a badge flaunted by those on the top.

Checkered History of Admissions Tests

Tests and college admissions have a century-long troubled history. Public universities, roughly between the 1890s and the late 1950s, used to admit everyone with a high school degree from a certified public high school. Then, in the 1950s, mid-western public universities developed the ACT as an alternative to the SAT; once the University of California, under Kerr's presidency, wanted to compete with Harvard, it signed up for the SAT in 1968, against the recommendations of every study produced by the university (see John Douglass' account in chapter 3 of *SAT Wars*), making standardized testing rather than high school grades the passkey to higher education. The direct link between public universities and public high schools was cut mid-century.

Private institutions launched the College Board in 1900 to set common exams on academic subjects that would give bragging rights to the private sector. Private colleges did not accept just any high school graduate, but only those who could do college level work in a particular subject as signified by their College Board exam scores. Then the Jewish community in New York blew by that academic hurdle, creating at Columbia University a Jewish foothold on the college/social mobility ladder. Fearing a Jewish invasion, anti-Semitic Yale and Princeton wanted an I.Q. test that would show, in the words of the Princeton psychologist who oversaw the design of the test, the superiority of their Nordic youths over inferior racial stock: the Alpine, Mediterranean (including Jews), and Negro (Soares, 2007). In the 1920s, I.Q. eugenics were not just an intellectual sub-culture, but rather the reverse - they were the law of the land with "separate but equal," forced "three generations of imbeciles are enough" sterilization, and strict immigration quotas. When the SAT was introduced in 1926 it was supposed to be an IQ test that would measure intrinsic intellectual aptitude, not academic subject mastery; it was supposed to help sort between the gems in the Nordic race from the subject-test grinds in the "Jewish race". It did not work to exclude Jews, but other tactics introduced in the 1930s of requiring mother's maiden name and place of birth, were more effective toward that goal. It also did not work to predict grades. Yale and Princeton knew that as early as 1930 (Soares, 2007). But the private sector clung to the test, first for the invidious distinction over public universities of requiring a national normed measuring stick, later because of the convenient way it disguised SES selection as academic selection, paying the bills along the way.

The lasting legacy was a pseudo-IQ test that sorted students by family income, opening or closing doors to colleges and careers in the process. We have traveled some considerable distance since then. The SAT's owners long ago discontinued using the name and the claim that it measured scholastic aptitude. Now the letters "SAT" do not reference anything, and the College Board only really claims the test predicts first year grades, which it does, but not well. There are significant defections even among the ranks of those who continue to embrace IQ bell curves. I take some considerable pleasure that Charles Murray, an author of the highly controversial *Bell Curve*, a man who believes firmly in the importance of IQ, joins me in calling for the abolition of the test. As Murray says in *SAT*

Wars, "The evidence has become overwhelming …. [S]o I find myself arguing that the SAT should be ended. Not just deemphasized, but no longer administered. … [T]he SAT score, intended as a signal flare for those at the bottom, has become a badge flaunted by those on the top" (Murray, 2012, p. 69). I also agree with Murray that the test will end when any of the top colleges, such as Harvard or Stanford, break with the farce. Murray wrote, "If just those two schools took such a step, many other schools would follow suit immediately, and the rest within a few years. …Admissions officers at elite schools are already familiar with the statistical story … They know that dropping the SAT would not hinder their selection decisions" (Murray, 2012, p. 80). It is high time for higher education to set aside the old golden compass, and to strike out for admissions tools worthy of the 21st century.

Test-Optional Admissions: Theory and Practice

Our non-test-score undergraduates perform academically as well as our test score submitters. We have not suffered any lowering of academic standards from the new policy; rather, there is considerable evidence of the reverse.

In *SAT Wars* there is a chapter jointly authored by two Princeton academics, a statistician, Chang Young-Chung, and sociology professor, Thomas Espenshade, which uses national data to model the impact on academic excellence and social diversity of doing admissions without relying on the SAT or ACT (Espenshade & Chung, 2012). Espenshade and Chung found results differed by type of institution. Private colleges were best served by going "test-optional." In their statistical simulation, private colleges got more racial and SES diverse and academically stronger students, as judged by high school grades and AP exam scores, by going test-optional. Public universities, on the other hand, did best by an admissions policy they dubbed "don't ask, don't tell," where the institution would not even look at test scores. State universities got academically stronger students, and more social diversity when they admit without any reference to test scores. It is a lesson reinforced by the findings on high school grades and standardized tests from the University of Georgia in chapter 8 of *SAT Wars* and by the findings of Bowen, Chingos, and McPherson's *Crossing the Finish Line: Completing College at America's Public Universities* (2009). Public universities waste taxpayers' money, distract students from focusing on learning the curriculum, and practice social discrimination when they require SAT/ACT scores.

In *SAT Wars*, I show how Wake Forest University's experience, now four years old, of conducting test-optional admissions has confirmed the statistical forecast offered by Espenshade and Chung (2012). In the academic year after the May 2008 announcement of Wake Forest's test-optional policy,

> Our applicant pool, even in the worse economic year in recent history, went up by 16%; our minority applicants went up by 70%. As reported in the *Journal of Blacks in Higher Education*, 6% of Wake Forest's senior cohort were minorities of color before the policy change; in the two [now three] cohorts admitted thus far as test-optional, the percentage of Black and Hispanic has gone up to 23. Asian student numbers have increased to 11%. First-generation youths, where neither parent went to college, jumped to 11%; Pell Grant youths, whose families earn near the poverty line, nearly doubled to 11%. In 2009, 78% of WFU undergraduates came from outside North Carolina (Soares, 2012a, p. 207).

Our academic strength has grown as well, as measured by entering students from the top ten percent of their high school class, which has gone up from 65 percent in 2008 to 83 percent in 2011 (Soares, 2012a). For research purposes and to monitor the test-optional policy, Wake Forest requires everyone admitted without a test score to send one before he or she arrives on campus. Accurate scores are reported to ratings publications, so no one can accuse the university of using this policy to artificially inflate our standing in the ratings game. Matriculating students are examined each semester to determine whether there are any differences between students who do or do not submit test scores. Wake Forest looks at course enrollment patterns, withdrawals from classes or from college, and grades achieved. As reported in detail in my conclusions in *SAT Wars*, we have found no statistically significant differences. Our non-test-score undergraduates perform academically as well as our test-score submitters. We have not suffered any lowering of academic

standards from the new policy; rather, there is considerable evidence of the reverse. We have found, along with the percent of our students from the top 10% of their high school classes having gone up dramatically, that library usage has increased as well. "Librarians are marvelous for keeping track of their domain, and from them we learned that library usage went way up: 63% increase in personal research sessions; 55% increase in instructional library sessions; 26% increase in credited library instructional classes; daily average visits went up by 10%; daily unique library web site visits went up by 62%" (Soares, 2012a, p. 209). Campus life, in and out of the classroom, looks and feels more diverse, more stimulating, and more engaging than ever before. When Wake Forest went test-optional, there were about 775 higher education institutions in that camp; today our ranks number 856. With nearly a third of all four-year degree granting institutions already with some form of test-optional admissions, the tipping point to push past the SAT/ACT is within sight.

References

Atkinson, R., & Geiser, S. (2012). Reflections on a century of college admissions tests. In J. A. Soares (Ed.), *SAT wars: The case for test-optional college admissions* (pp. 23-49). New York, NY: Teachers College Press.

Bowen, W., Chingos, M., & McPherson, M. (2009). *Crossing the finish line: Completing college at America's public universities.* Princeton, NJ: Princeton University Press.

Espenshade, T., & Chung, C. Y. (2012). Diversity outcomes of test-optional policies. In J. Soares (Ed.), *SAT wars: The case for test-optional college admissions* (pp. 177-200). New York, NY: Teachers College Press.

Geiser, S., & Santelices, M. V. (2007). *Validity of high-school grades in predicting student success beyond the freshman year: High-school record vs. standardized tests as indicator of four-year college outcomes* (Occasional Paper Series, 6.07). Berkeley, CA: Center for Studies in Higher Education.

Kobrin, J., Patterson, B., Shaw, E., Mattern, K., & Barbuti, S. (2008). *Validity of SAT for predicting first year college grade point average* (Report No. 2008-5). New York, NY: College Board.

Morgan, R. (1989). *Analyses of the predictive validity of the SAT and high school grades from 1976 to 1985* (Report No. 89-7). New York, NY: College Board.

Murray, C. (2012). Abolish the SAT. In J. Soares (Ed.), *SAT wars: The case for test-optional college admissions* (pp. 69-81). New York, NY: Teachers College Press.

Rosner, J. (2012). The SAT: Quantifying the unfairness behind the bubbles. In J. Soares (Ed.), *SAT wars: The case for test-optional college admissions* (pp. 104-117). New York, NY: Teachers College Press.

Soares, J. A. (2007). *The power of privilege: Yale and America's elite colleges.* Stanford, CA: Stanford University Press.

Soares, J. A. (Ed.). (2012a). *SAT wars: The case for test-optional college admissions.* New York, NY: Teachers College Press.

Soares, J. A. (2012b). The future of college admissions: Discussion. *Educational Psychologist, 47*(1), 66-70.

Sternberg, R.J., Bonney, C., Gabora. L., & Merrifield, M. (2012). WICS: A model for college and university admissions. *Educational Psychologist, 47*(1), 30-41.

Note: The Golden Compass, NY, Knopf, 2002; The Subtle Knife, NY, Knopf, 2001; The Amber Spyglass, NY, Knopf, 2003.

Abstract

Despite the extensive testing for federal accountability mandates, college students' understanding of federal accountability testing (e.g., No Child Left Behind, Race to the Top, Spellings) has not been examined, resulting in a lack of knowledge regarding how such understanding (or lack thereof) impacts college students' behavior on accountability tests in higher education contexts. This study explores college students' understanding and misconceptions of federal accountability testing in K-12. To this end, we crafted nine multiple choice items with four distracters and piloted these items with two college student samples. The results indicated that college students tend to be moderately confident in their responses regardless of the accuracy of the response. These findings imply that educating students on the purpose and process of accountability testing will require not only imparting correct information, but also debunking misconceptions.

AUTHORS

Anna Zilberberg, M.A.
James Madison University

Robin D. Anderson, Psy.D.
James Madison University

Peter J. Swerdzewski, Ph.D.
Regents Research Fund

Sara J. Finney, Ph.D.
James Madison University

Kimberly R. Marsh, M.A.
James Madison University

CORRESPONDENCE

Email
zilberax@jmu.edu

Growing Up with No Child Left Behind: An Initial Assessment of the Understanding of College Students' Knowledge of Accountability Testing

"Many of the criticisms we hear about educational assessments appear to be based on misconceptions. Some of them are due to persons simply misunderstanding the meaning of test scores and their implications for instructional improvement and school accountability" (Goodman & Hambleton, 2005, p. 107).

Accountability testing in educational settings has been on the increase due to federal mandates, such as No Child Left Behind (NCLB, 2002), Race to the Top-related testing initiatives (Obama, 2009), and the Spellings report (2006). Despite the widespread use of accountability testing, little is known about students' understanding of accountability testing, and even less is known about how this understanding (or lack thereof) impacts students' test-taking behavior (e.g., effort, honesty). For example, do students who understand how K-12 accountability test results are used give more test-taking effort on the accountability assessments they complete in college than those students who do not know how K-12 accountability test scores are used? If students understand the role of the federal government in the K-12 accountability process, are they more or less likely to give their best effort on accountability tests they encounter in college? These are all empirical questions. However, before answering these questions, a more fundamental question must be answered–do students understand accountability testing mandates at all? It may be the case that students have very limited understanding of these mandates. On the other hand, given their extensive experience of being tested in K-12, they may have learned the purposes behind the testing process.

Of course, whether or not students possess an understanding of K-12 testing mandates is an empirical question.

The purpose of the current study is to provide an initial assessment of college students' understanding of testing associated with federal K-12 institutional accountability mandates (e.g., NCLB)—testing these students experienced for numerous years. That is, although the current generation of college students has experienced accountability testing from elementary school to college, little is known about how well students understand K-12 accountability testing, and how this understanding, or lack thereof, impacts students' test-taking behavior on the accountability tests they complete in college. An item-by-item examination of responses to carefully crafted items representing key aspects of accountability testing provided insight into college student misconceptions regarding such testing. Furthermore, by examining the confidence students have in the correctness of their responses, we begin to understand how difficult it might be to change these misconceptions. Before presenting our findings, we first emphasize the importance of examining college students' knowledge of institutional accountability testing in K-12 and review the literature in this domain.

Misconceptions about Institutional Accountability Testing

So, what are the core concepts of accountability testing imperative for students (and teachers and the public at large) to know? Sireci (2005) discusses six fundamental concepts about assessment. A basic understanding of these concepts is necessary for forming "intelligent opinions about the quality and appropriateness of tests" (Sireci, 2005, p. 112). These concepts are: (a) what is a standardized test; (b) the difference between norm-referenced and criterion-referenced tests; (c) reliability; (d) validity; (e) the setting of passing test scores; (f) obtaining more information about the test (e.g., where to find information about the test development process). Basic understanding of these concepts is a necessary precursor to a critical evaluation of the worth of accountability testing. However, Sireci noted that many criticize tests without adequate background knowledge of these critical concepts that underpin the testing process.

Criticisms voiced against accountability testing include narrowed curriculum, allocation of valuable instructional time toward testing and test preparation, high costs, increased cheating, over-reliance on a single test score, and biased test items (Goodman & Hambleton, 2005; Ravitch, 2010; Sireci, 2005). These criticisms spark debates in the arena of educational policy. Many educational professionals (e.g., teachers, administrators) question whether accountability programs actually serve to improve the quality of education (e.g., Abrams, Pedulla, & Madaus, 2003; Jones et al., 1999). However, there may be fundamental problems pertaining to the sources of many test-related criticisms if these criticisms are due to a lack of understanding of psychometric and policy-related concepts. Although some of the criticisms mentioned above embody legitimate concerns, many may be based on misconceptions about testing.

Despite the wide-spread use of accountability testing, little is known about students' understanding of accountability testing, and even less is known about how this understanding impacts students' test-taking behavior.

If students are misinformed about the basics of testing, that would imply students believe they have some general knowledge about the fundamental assessment-related concepts outlined by Sireci (2005), but in fact that knowledge is, at least to some extent, inaccurate. Thus, altering negative attitudes about accountability testing entails not only imparting accurate knowledge, but also identifying and debunking misconceptions. Leading researchers in the field of psychometrics outline some of these misconceptions. Goodman and Hambleton (2005) draw from their experience in the field of psychometrics when discussing several assessment-related misconceptions that are due to "misunderstanding the meaning of test scores and their implications for instructional improvement and school accountability" (p.107). These authors identified four misconceptions held by the general public: (a) high-stakes assessments set everyone involved up for a failure; (b) a single test score is used to make high-stakes decisions; (c) test items are biased; and (d) performance standards are set too high. Although the anecdotal evidence pertaining to the general lack of knowledge and misconceptions about accountability testing is overwhelming and infor-

mative, empirical questions about how these misconceptions are related to attitudes toward tests, test-taking effort, and test performance cannot be addressed without a measure of student knowledge of accountability testing.

If We Care About Student Knowledge and Misconceptions, How Do We Assess It?

No measure of student knowledge of K-12 accountability testing currently exists. This is not surprising as it would be difficult to create given the breadth of the construct. This study is the first attempt to assess students' knowledge and should be viewed as such – a pilot study that provides initial insight into students' understanding and misunderstanding of accountability testing mandates. We created a set of items to address three aspects of students' knowledge. First, we were interested in investigating the extent to which students are aware of what exactly is mandated in terms of academic achievement in public schools (e.g., what "proficiency" entails in this context and what the test results are used for). Second, we were interested in learning whether students held any misconceptions in regard to how state-mandated testing was carried out in schools (e.g., what percentage of the academic year is taken up by testing). Third, we intended to learn whether students knew who the different stake-holders involved in state-mandated testing were and their respective roles (e.g., who sets the standards?).

…By examining the confidence students have in the correctness of their responses, we begin to understand how difficult it might be to change these misconceptions.

As higher education assessment practitioners, we believe that understanding college students' misconceptions about K-12 accountability testing is valuable in understanding college students' perceptions of higher education assessment testing. While we acknowledge that there are presently no specific nationally mandated tests for college students, increasing demands for accountability by the federal government essentially translates to mandates via the standards set by regional accreditors. This results in a K-16 continuum of assessment "mandates". Failure to explore the possible impact of K-12 testing on college students' performance on higher education assessments could result in inappropriate inferences regarding students' progress and program effectiveness. For this reason, we are focusing on college students' understanding of K-12 accountability assessment.

Domains of Student Knowledge of Accountability Testing

Given the variation with which states implement state-mandated policy on testing, most investigations of accountability testing focus on a single state (e.g., Jones et al., 1999). Despite the many nuances in how states enact calls for accountability, several federal provisions apply equally to all states. In other words, there are common features pertaining to implementation of NCLB across all states. For this reason, it makes sense to examine students' knowledge of the basic premise underlying institutional accountability testing in K-12; that is, the purpose and the intended use of these federally mandated tests.

In order to develop a set of multiple choice items used in this study, a team of subject matter experts employed a careful and systematic approach. The team consisted of two faculty members with extensive expertise in psychometrics and higher education assessment policy and accountability issues, two advanced doctoral students in Assessment and Measurement, and a content expert in K-12 accountability issues. Combined, the team has thirty years of experience in assessment, accountability testing, and instrument design.

We constructed nine multiple-choice items to address the key aspects of NCLB universally applicable to all states. The content expert in K-12 accountability issues reviewed relevant literature and identified key aspects of NCLB applicable across states. The following key aspects were subsumed under the "What" category of K-12 accountability testing: Schools must experience growth (called Annual Yearly Progress, or AYP) toward proficiency each year; academic proficiency at each level is defined by the state; and the goal of NCLB is adequate education for all. The following key aspects were subsumed under the "Who" category: The federal government administers penalties to schools that fail to achieve proficiency; and the state government sets the learning standards with which the account-

ability tests must align. The following key aspects were subsumed under the "How" category: Detailed information of performance of each school and each of the four subgroups must be publicly available and readily accessible via the school's report card, which must be provided to parents; NCLB states that school accountability is based only on student performance; factors such as resources, classroom sizes, parent involvement, etc. are not considered; and on average, students spend only about 1% of their total school year taking NCLB-required tests. Several concepts initially identified by the content expert were ruled out during the review process because they were deemed to be too specific or too advanced for students to know. For example, an interesting aspect of NCLB is that the federal act only disciplines schools and districts for poor student performance; whether or not an individual teacher is disciplined due to poor performance on NCLB tests is a state issue. Although a noteworthy side of NCLB, a multiple choice item was not created to address this aspect specifically. The process of delineating these key aspects listed above marked the beginning of an iterative item creation process. Next, the team members carefully crafted and reviewed the stems, distractors, and correct responses for each one of the multiple choice items. The resulting nine items are a product of this iterative and systematic process of item development. Nonetheless, we must stress that we do not assume that the sum of these items represent one test of a unidimensional construct of knowledge of accountability testing. Instead, these nine items allow initial insight into students' misconceptions about specific testing issues.

In order to gauge the degree of confidence that students possess with respect to their answers, a Likert-type item, prompting students to rate their level of confidence in their response to each knowledge item, was included after each one of the multiple-choice items. We were interested in how strongly students held their misconceptions regarding accountability testing, as strongly held misconceptions may be more difficult to correct than those held with less confidence. That is, it was of interest to gauge the strength of students' beliefs in the accuracy of their knowledge. The items can be found in the Appendix.

…Altering negative attitudes about accountability testing entails not only imparting accurate knowledge, but also identifying and debunking misconceptions.

Methods

The items were administered as part of a large-scale university assessment effort at a mid-sized, mid-Atlantic four-year institution. Two samples of college students completed the items: (a) incoming college freshmen, and (b) college sophomores.

A total of 3606 incoming college freshmen were administered the items the summer before attending college. A total of 3196 attempted all nine items, thus this sample serves as the sample under study. Females comprised 62.47% of the sample, with less than 1% of students not indicating their gender. About 9.5% were 17 years of age or younger, 85.5% were 18, 4.5% were 19, less than 1% were over 20, with less than half-a-percent choosing not to indicate their age. Of the 3191 students who reported their ethnicity, 85% were Caucasian, 5.6% were Asian, 3.5% were African-American, 2.4% were Latino, and 5.75% were Native-American or multiracial. The majority of students (70%) were from Virginia, 29% were from outside of Virginia, and less than 1% were from outside of the United States, with about 0.59% of students choosing not to indicate their geographic area. Most students reported their high school GPA to be A- or above (59.79%), followed by B- and above (40.12%); less than 1% of students reported a GPA of C+ and below.

The sophomore student sample consisted of 424 students who were administered the items as part of a university-wide assessment day. A total of 382 attempted all nine items. Demographic information was available for 380 participants: 62.8% were female, 77% were Caucasian, average age was 19.15 (SD=0.88), average GPA was 3.05 (SD=0.56), and less than 1% were 18 or younger.

Results

Analyses were conducted at the item level, providing specific information regarding knowledge and confidence of distinct aspects measured by each one of the items. The distinctiveness of the nine items is empirically supported via weak relationships among these

diverse items: correlations ranged from -0.07 to 0.39 for freshmen and from -0.111 to 0.315 for sophomores.

Descriptive statistics for the items and corresponding confidence items for both freshman and sophomore samples can be found in Table 1. Overall, the items ranged in difficulty from 0.20 to 0.74 for the freshman sample and from 0.20 to 0.74 for the sophomore sample. Next, we examined students' responses to the items to highlight students' misconceptions. The item results are organized by domain of knowledge (i.e., the "what", "who", and "how" of accountability testing).

Table 1

Descriptive Statistics for Nine Multiple-Choice and Confidence Items for Freshman and Sophomore Samples

	Freshman Sample ($N = 3196$)				Sophomore Sample ($N = 382$)			
	Knowledge Items		Confidence Items		Knowledge Items		Confidence Items	
Item	Item Difficulty	*SD*	Mean	*SD*	Item Difficulty	*SD*	Mean	*SD*
1	0.24	0.43	5.07	1.27	0.36	0.48	4.90	1.32
2	0.25	0.43	4.01	1.51	0.26	0.44	3.98	1.42
3	0.56	0.50	4.95	1.33	0.50	0.50	4.91	1.34
4	0.28	0.45	4.70	1.32	0.28	0.45	4.66	1.36
5	0.37	0.48	4.94	1.27	0.45	0.50	4.70	1.25
6	0.48	0.50	3.60	1.58	0.55	0.50	3.70	1.52
7	0.20	0.40	3.80	1.38	0.20	0.40	3.73	1.37
8	0.62	0.48	4.01	1.64	0.67	0.47	3.70	1.66
9	0.74	0.44	4.75	1.40	0.74	0.44	4.60	1.42

Note. Confidence items reflect students' confidence in their response on each one of the items; higher scores indicate a greater degree of confidence (Likert-type scale ranging from 1 - *not confident*, 4 – *moderately confident*, 7 – *completely confident*).

"What"

Failure to explore the possible impact of K-12 testing on college students' performance on higher education assessments could result in inappropriate inferences regarding students' progress and program effectiveness.

For both the freshman and sophomore students, less than 45% of the students answered each of the three "what" items correctly. Furthermore, two of the three items were answered correctly at a guessing rate of 0.25 or very close to it, supporting our conclusion that students do not know the correct responses to these items.

What: Item 1. Item 1 evaluated whether students could correctly identify the goal of institutional accountability testing, with the correct response being that testing is used to determine if a given student is on-track for proficiency. Approximately 24% (below the guessing rate) of freshmen answered the item correctly, whereas 36% (SD = 0.48) of the sophomores answered this item correctly (see Table 2). Most students incorrectly endorsed the response option indicating that the most important goal of the state is to ensure that "every student answer enough questions correctly to indicate the student is proficient in the subject every year" (freshmen = 72%; sophomores = 61%).

What: Item 4. Item 4 examined students' knowledge of the purpose of the NCLB Act (which is to ensure adequate access to education for all students). About 28% of freshmen and sophomores answered this item correctly (just a few percentage points above the guessing rate). Notably, the majority of freshmen (55%) and sophomores (58%) endorsed the incorrect response option indicating that the act is specifically designed to ensure that all students in the United States are meeting the same national standards of learning in academic areas including math, reading, and science.

What: Item 5. Item 5 evaluated students' knowledge of what is meant by proficiency as operationalized by the state-mandated tests (i.e., mastery of grade-level work as defined by state). Approximately 37% of freshmen correctly responded to the item and about 45% of sophomores answered this item correctly. However, more students endorsed

the incorrect response option that defined proficiency as having enough knowledge and skill to be successful in the next grade level (freshmen = 53% and sophomores = 51%).

"Who"

For both the freshman and sophomore students, the percentage of students answering each of the three "who" items correctly varied widely (e.g., 26% of students answering correctly on one item compared to 74% answering correctly on another). One of the three items was answered at a guessing rate or close to it, suggesting that students most likely do not know the correct response to that item.

Who: Item 2. Item 2 examined students' knowledge of the repercussions associated with students not performing well on the tests, with the correct response option being that schools are penalized in various ways. About 25% of freshmen answered the item correctly (just at the guessing rate); likewise, approximately 26% (just above the guessing rate) of the sophomores answered this item correctly. The most frequently endorsed option was the incorrect response that students get held back a grade until the student learns enough to pass the test (47% of freshmen and 41% of sophomores endorsed this option). This finding reflects students' confusion regarding the federal mandates versus the implementation of these mandates in certain states and districts.

Who: Item 3. Item 3 examined students' knowledge regarding who sets the standards for the state-mandated tests, with the correct response option being that specific standards are set by the state. About 56% of freshmen selected the correct answer and about 50% of sophomores answered this item correctly. The most frequently selected incorrect answer among freshmen (36%) and sophomores (45.8%) was that the U.S. Department of Education is the standard-setting body.

Who: Item 9. Item 9 evaluated students' knowledge regarding which governing body selects the specific content for federal accountability tests, with the correct response option being that content is set by the state. About 74% of freshmen and sophomores answered this item correctly. The second-most-frequently endorsed answer was the incorrect response option suggesting that the federal government is responsible for the specific content on federal accountability tests (19% of freshmen and 21% of sophomores endorsed this option).

"How"

For both the freshman and sophomore students, the majority of students answered only one of the three "how" items correctly. One of the three items was answered below the guessing rate, further suggesting that students most likely do not know the correct response to that item.

How: Item 6. Item 6 examined students' knowledge of the reporting requirements on a school's "report card" (i.e., report average scores by grade and by ethnic group). About 48% of freshmen answered the item correctly, whereas approximately 55% of sophomores answered this item correctly. The second-most-frequently endorsed answer was the incorrect response that the individual scores, with names concealed, are listed on the "report card" (29% freshmen and sophomores endorsed this option).

How: Item 7. Item 7 evaluated students' knowledge of factors used to evaluate the effectiveness of schools, with the correct response option being that test scores are the only factor used for the purposes of federal accountability. Only about 20% of freshmen and sophomore students answered this item correctly (below the guessing rate). The three distracters, which focused on financial resources, SES of students, and school size and location, were almost equally endorsed. This was true for both the freshman and sophomore samples.

How: Item 8. Item 8 examined students' knowledge regarding the average amount of time students spend annually taking state-mandated tests, with the correct response option being that about 1% of the academic year is used for the administration of federal accountability tests. About 62% of freshmen answered the item correctly, whereas about

67% of sophomores answered this item correctly. The most-often-endorsed incorrect response indicated that 7% of the school year is devoted to testing (26% of freshmen and 23% of sophomores endorsed this option).

Table 2
Percent of Students Endorsing Each of the Nine Items by Area for Freshman and Sophomore Samples

Area	Test Items	Percent	
		Freshmen: (N = 3196)	Sophomores (N = 382)
WHAT	Item 1		
	Response (a)	3.63	2.62
	Response (b)	0.59	0.79
	Response (c)	72.06	60.99
	Response (d)*	23.72	35.60
	Item 4		
	Response (a)	2.69	1.05
	Response (b)	55.48	57.59
	Response (c)	13.86	13.61
	Response (d)*	27.97	27.75
	Item 5		
	Response (a)	53.41	50.79
	Response (b)	4.32	1.31
	Response (c)	5.41	3.4
	Response (d)*	36.86	44.5
WHO	Item 2		
	Response (a)	4.1	3.4
	Response (b)	46.53	40.84
	Response (c)	24.19	29.84
	Response (d)*	25.19	25.92
	Item 3		
	Response (a)	2.94	2.09
	Response (b)	4.88	1.83
	Response (c)*	55.73	50.26
	Response (d)	36.45	45.81
	Item 9		
	Response (a)	0.81	0.79
	Response (b)	5.73	3.4
	Response (c)*	74.28	74.35
	Response (d)	19.18	21.47
HOW	Item 6		
	Response (a)	13.61	10.99
	Response (b)	28.91	28.8
	Response (c)	9.89	4.97
	Response (d)*	47.59	55.24
	Item 7		
	Response (a)	23.81	26.18
	Response (b)	24.69	22.51
	Response (c)	31.07	31.15
	Response (d)*	20.43	20.16
	Item 8		
	Response (a)*	62.27	66.75
	Response (b)	26.1	23.3
	Response (c)	8.48	7.85
	Response(d)	3.16	2.09

Note. *indicates correct response

Students' Confidence in the Accuracy of their Responses

Recall that both freshman and sophomore students were asked to rate the confidence level they had in the accuracy of their responses. Even though students from both samples answered most of the items incorrectly, their confidence ratings reflected that they were moderately confident in the accuracy of their responses (on a 7-point scale with a value of 4 representing *moderately confident*, average confidence ratings for items ranged from 3.60 to 5.07 for freshmen and from 3.70 to 4.91 for sophomores). Upon examining

student confidence separately for those who responded to the item correctly versus those who responded incorrectly, we noted several trends across the two samples (see Table 3). First, effect size estimates (Cohen's *d*) indicated that there were negligible differences between student confidence ratings from those who responded correctly versus from those who responded incorrectly to items 1, 2, 5, and 7. In other words, although we would hope that those responding correctly would have more confidence in their response than those responding incorrectly, that did not occur for four of the items. Importantly, these four items did not represent the same domain (e.g.,"What") but instead cut across each of the aspects of accountability testing (with items 1 and 5 corresponding to the "What", item 2 corresponding to the "Who", and item 7 corresponding to the "How" domains of accountability testing, respectively).

Not only did they not know the basic premises underlying federal institutional accountability testing in K-12, but they also believed that their misconceptions were correct.

Second, for items 6 (factors used to evaluate the effectiveness of schools), 8 (the amount of time spent on the administration of federal accountability tests during the year), and 9 (what the tests are designed to measure), freshman and sophomore students who responded correctly were significantly more confident in the accuracy of their response than those who responded incorrectly. It should be noted that the largest effect size estimates were consistently observed for items 6 and 8 across both samples. These two items correspond to the "How" domain of accountability testing, ultimately indicating that students may be more accurate in their appraisal of their knowledge in relation to this specific domain.

Table 3
Differences in Average Confidence Ratings between Students Who Answered Each Item Correctly and Incorrectly

Freshmen ($N = 3196$)

	Correct			Incorrect						Cohen's
Item	*Mean*	*SD*	*N*	*Mean*	*SD*	*N*	Difference	*t*	*p*	*d*
1	4.88	1.30	758	5.12	1.25	2438	0.25*	4.67	<0.01	0.19
2	4.09	1.58	805	3.98	1.49	2391	-0.11	1.69	0.09	-0.07
3	5.08	1.26	1781	4.78	1.40	1415	-0.30*	6.33	<0.01	-0.23
4	4.61	1.32	894	4.74	1.32	2302	0.13	2.50	0.01	0.10
5	5.03	1.26	1707	4.83	1.28	1489	-0.20*	4.45	<0.01	-0.16
6	3.92	1.61	1521	3.31	1.50	1675	-0.61*	11.07	<0.01	-0.39
7	3.86	1.44	653	3.79	1.36	2543	-0.07	1.14	0.26	-0.05
8	4.38	1.63	1990	3.41	1.47	1206	-0.97*	17.31	<0.01	-0.62
9	4.89	1.34	2374	4.35	1.50	822	-0.53*	9.01	<0.01	-0.39

Sophomores ($N = 382$)

	Correct			Incorrect						Cohen's
Item	*Mean*	*SD*	*N*	*Mean*	*SD*	*N*	Difference	*t*	*p*	*d*
1	4.82	1.14	136	4.95	1.41	246	0.12	0.88	0.38	0.09
2	3.93	1.47	99	3.99	1.40	283	0.06	0.38	0.70	0.05
3	4.92	1.25	192	4.89	1.42	190	-0.03	0.24	0.81	-0.02
4	4.41	1.23	106	4.76	1.40	276	0.36	2.30	0.02	0.26
5	4.78	1.18	170	4.63	1.31	212	-0.14	1.12	0.26	-0.12
6	3.95	1.62	211	3.4	1.33	171	-0.54*	3.53	0<.01	-0.36
7	3.68	1.53	77	3.74	1.34	305	0.07	0.37	0.71	0.05
8	4.02	1.66	255	3.05	1.48	127	-0.97*	5.60	<0.01	-0.61
9	4.69	1.38	284	4.34	1.49	98	-0.35	2.14	0.03	-0.25

Note. Confidence items reflect students' confidence in their response on each one of the multiple-choice items; higher scores indicate a greater degree of confidence (Likert-type scale ranging from 1 - *not confident*, 4 - *moderately confident*, 7 - *completely confident*). Positive difference indicates that those who answered the item incorrectly were more confident in their response than those who answered the item correctly. An asterisk (*) signifies statistical significance at the $p < 0.01$ level. Cohen's *d,* a practical significance measure of the magnitude of an observed difference between two means on a standardized metric, (Cohen, 1988) is calculated based on pooled standard deviation.

Third, freshman and sophomore samples differed in their confidence rating patterns for only two items: items 3 (who deems what students are supposed to learn and what test content is aligned to) and 4 (the overall purpose of NCLB). Freshmen responding correctly to item 3 indicated significantly higher levels of confidence in the accuracy of their response in comparison to freshmen that responded incorrectly. Conversely, there were negligible differences between sophomore student confidence ratings from those who responded

correctly versus incorrectly to item 3. For item 4, there were negligible differences between freshman student confidence ratings from those who responded correctly versus incorrectly. Interestingly, this was not the case for sophomore students, with students responding incorrectly to item 4 indicating significantly higher levels of confidence in the accuracy of their response in comparison to those who responded correctly. Overall, students' confidence ratings indicate strongly held misconceptions regarding accountability testing.

Discussion

The purpose of the current study was to provide an initial assessment of college students' understanding of K-12 accountability mandates. The nine items piloted in this study were carefully crafted to assess college students' knowledge of, and misconceptions about, K-12 institutional accountability testing associated with NCLB mandates. Specifically, the following three aspects of accountability testing were addressed: what such tests entail, how the results are used, and who mandates testing. In addition, a Likert-type scale confidence item accompanied each of the multiple choice items to allow for the measurement of the degree of confidence that students had in the accuracy of their responses.

Results pertaining to both knowledge (i.e., correctness of response) and confidence followed a similar pattern for both freshman and sophomore samples. More specifically, students hold misconceptions in all three areas addressed by the items: the "what", "who", and "how" of accountability testing. Pertaining to "what", the majority of freshmen and sophomores erroneously believe that the purpose of NCLB is to impose national standards of learning (as opposed to providing equal access to adequate education as defined by the state), to define proficiency as being successful at the next level (as opposed to staying on track to proficiency). Pertaining to "who", a majority of students erroneously believe that the federal government holds students back a grade if test results do not meet the standards (as opposed to schools receiving various penalties); approximately half of all students believe that the U.S. Department of Education sets the standards (as opposed to state), and about 20% of students believe that the federal government (as opposed to state) selects the content to be covered on the tests. Pertaining to "how", only about half of the students know that the school "report card" includes average scores broken down by ethnic group (about 29% of all students think that individual student scores are reported) and both freshmen and sophomores hold misconceptions as to what factors are used for evaluating school effectiveness. Fortunately, the majority of students in both samples know how much time is devoted to federal accountability test administration (i.e., about 1% of the academic year). On average, both freshmen and sophomores hold common misconceptions regarding institutional accountability testing.

In addition, it appears that students tend to confuse the actual state mandates with the schools' practice or implementation. For example, students' responses to items 2 and 5 illustrate students' experience of needing to pass the accountability test in order to advance to the next grade. Although the federal mandate does not require individual students to pass the test in order to advance to the next grade level, many states and districts do impose this requirement. Thus, students assume that the passing requirement is due to the mandate, whereas in reality it is due to the state- or district-specific implementation of the mandate.

Educating students about the purposes of assessment might result in more accurate test scores.

Evaluation of students' confidence levels in their responses reveals that students are confident in these beliefs even when the students are wrong. On average, both freshmen and sophomores were moderately confident in their responses (with variability being slightly higher in the freshman sample), even though both groups of students answered the majority of the items incorrectly. In other words, students' evaluation of their own knowledge was inaccurate; it was biased upward. That is, judging by the students' self-reported moderate confidence in their erroneous responses, not only did they not know the basic premises underlying federal institutional accountability testing in K-12, but they also believed that their misconceptions were correct.

The current study is not free of limitations, but there are several ways in which future research can remedy these limitations and build on our findings. The purpose of this

preliminary investigation was to examine students' understanding of several key aspects of K-12 accountability, as well as identify the misconceptions most common among students, as represented by the distracters. Distracter analyses not only allowed the researchers to identify gaps in student knowledge of accountability testing, but also students' common misconceptions. Future instrument development studies should continue to give careful attention to the distracters. Furthermore, an examination of test-retest reliability should be conducted in order to garner evidence for the stability of scores across administrations. In addition, future item construction initiatives would benefit from item reviews conducted by independent content experts. Even though the current study is not a full scale development study, we believe that this initial investigation sets the stage for such future research endeavors.

Implications and Conclusions

In the context of accountability testing, it might be the case that college students who are ill-informed about what K-12 accountability tests entail, how the results are used, and who is mandating the tests are more likely to develop negative attitudes toward all large-scale accountability testing, are less likely to alter their attitudes toward such tests, and are therefore less likely to exert effort on the accountability tests they complete both in K-12 and college, jeopardizing the validity of inferences made based on these test scores. That is, educating students about the purpose of assessment might result in more accurate test scores. For example, Huffman, Adamopoulos, Murdock, McDermid, and Cole (2011) found that college students who were exposed to an informative motivational presentation about the purpose of a program-level assessment scored higher on average on this assessment than students who received a monetary incentive, or those students who received no treatment (no presentation, no money). Given this finding, it is not surprising that others have called for informing students about the purpose of accountability testing (Leveille, 2006; Zilberberg, Brown, Harmes, & Anderson, 2009).

One may assume that simply educating students about the mandates will resolve the issue of undesirable attitudes and allow students to form appropriate attitudes based on accurate information, which would subsequently improve test-taking behavior. However, in addition to the concern that students may lack knowledge about accountability testing, there is the equally worrisome concern that students may falsely believe they understand the core concepts of accountability assessment, when in reality their understanding is flawed and based on misconceptions. In other words, the issue at hand is more complicated if students are not merely uninformed about these concepts, but instead misinformed.

It is critical that states seeking the waivers and the federal government granting the waivers understand what students know about accountability testing and – just as importantly – what students misunderstand about accountability testing.

It follows that if one's goal was to educate students on the basics of accountability testing so that students can hold well-informed, intelligent opinions and develop appropriate attitudes, the challenge will not be just imparting knowledge; educational intervention will also entail debunking pre-existing misconceptions and shattering students' ungrounded confidence. As American historian Boorstin noted, "The greatest obstacle to discovery is not ignorance - it is the illusion of knowledge". Keeping this challenge in mind, future research endeavors can focus on developing a measure and using it for designing and evaluating such educational interventions. No time is more important than the present. Just as this article is being submitted for publication, President Obama has begun to implement the NCLB waiver program (McNeil & Klein, 2011). As states consider tailored plans for accountability that comply with the NCLB waiver requirements, it is critical that states seeking the waivers and the federal government granting the waivers understand what students know about accountability testing and—just as importantly—what students misunderstand about accountability testing. The successful adoption and implementation of revised accountability structures are predicated on knowing what students know about accountability testing, and understanding what actions students take based on this knowledge. As educational policy changes so must the multiple choice items assessing students' understanding of this educational policy.

Importantly, the current findings may also be relevant for developing a measure of knowledge of accountability testing in higher education. College students may differ in

what they know about accountability testing in K-12 versus accountability testing in college. Moreover, the relationships between such knowledge (or lack thereof) and test-taking behavior (e.g., effort, honesty) may differ depending on the context. That is, knowledge regarding accountability testing in higher education may have a stronger relationship with test-taking behavior on higher education accountability testing than knowledge of K-12 accountability testing. It would also be interesting to assess if knowledge of K-12 accountability testing is related to knowledge of higher education accountability testing. To answer these empirical questions, a higher education version of the items is needed.

In closing, we believe that the results of this pilot study provide an initial assessment of college students' understanding of accountability testing in K-12. As a preliminary investigation of the construct not previously researched, this study sets the stage for future full-scale test development studies, which would entail independent content review of the items as well as gathering reliability and validity evidence for the measure. After a reliable and valid method for measuring students' understanding of K-12 accountability is developed, numerous empirical questions can be addressed, such as the relationship between students' knowledge of accountability testing, students' attitudes toward such tests, and students' test-taking effort.

References

Abrams, L. M., Pedulla, J. L., & Madaus, G. F. (2003). Views from the classroom: Teachers' opinions of statewide testing programs. *Theory into Practice, 42*(1), 18-29.

Ajzen, I., & Fishbein, M. (1977). Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin, 84*, 888-918.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research.* Reading, MA: Addison-Wesley.

Goodman, D., & Hambleton, R. K. (2005). Some misconceptions about large-scale educational assessments. In R. P. Phelps (Ed.), *Defending standardized testing* (pp. 91-110). Mahwah, NJ: Lawrence Erlbaum Associates.

Huffman, L., Adamopoulos, A., Murdock, G., Cole, A., & McDermid, R. (2011). Strategies to motivate students for program assessment. *Educational Assessment*, *16*(2), 90-103.

Jones, G., Jones, B., Hardin, B., Chapman, L., Yarbrough, T., & Davis, M. (1999).The impacts of high-stakes testing on teachers and students in North Carolina. *Phi Delta Kappan*, *81*(3), 199-203.

Leveille, D. E. (2006). Accountability in higher education: A public agenda for trust and cultural change. Berkeley: Center for Studies in Higher Education.

Linn, R. L. (2000). Assessments and accountability. *Educational Researcher, 23*(9), 4-16.

McNeil, M., & Klein, A. (2011, September 28). Obama outlines NCLB flexibility; plan waives cornerstone provisions of law. *Education Week*, pp. 1-3.

No Child Left Behind Act of 2001, 20 U.S.C.S. Section 6301 et seq. (2002). Retrieved March 2nd, 2011, from http://www.ed.gov/nclb/overview/importance/difference/index.html

Obama, B. (2009, July 23). Interview by M.D. Shear & N. Anderson [Audio recording].President Obama Discusses New 'Race to the Top' Program. Washington Post Education Archives, Washington, DC. Retrieved from http://www.washingtonpost.com/

Petty, R. E., & Wegener, D. T. (1998). Attitude change: Multiple roles for persuasion variables. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Handbook of Social Psychology,* 4th ed., vol. II (pp.323-390). New York, NY: McGraw-Hill.

Ravitch, D. (2010). *The death and life of the great American school system: How testing and choice are undermining education.* New York, NY: Basic Books.

Sireci, S.G. (2005). The most frequently unasked questions about testing. In R. Phelps (Ed.), *Defending standardized testing* (pp. 111-121). Mahwah, NJ: Lawrence Erlbaum.

Spellings, M. (2006). *A test of leadership: Charting the future of U.S. higher education.* A report of the commission appointed by the Secretary of Education. Washington DC: U.S. Department of Education.

Zilberberg, A., Brown, A. R., Harmes, J. C., & Anderson, R. D. (2009). How can we increase student motivation during low-stakes testing? Understanding the student perspective. In D. McInerney, G. Brown, & G. A. Liem (Eds.), *Research on Sociocultural Influences on Motivation and Learning* (Vol. IX). Greenwich, CT: Information Age Press.

Appendix

Nine Multiple-Choice Items

Directions: Below are a series of questions designed to examine your understanding of state-mandated tests – tests that students must take in public elementary, middle, and high school (For example, in Virginia these tests are called Standards of Learning (SOL); we are not referring to tests such as SAT, PSAT, or ACT).
After selecting your response to each of the multiple choice items, please rate the level of confidence in your response.

1. For the state-mandated tests that students must take in public elementary, middle, and high school, the most important goal for the state is:

(a) For every student to answer every question correctly every year.
(b) For those students who are going to college to answer every question correctly every year.
(c) For every student to answer enough questions correctly to indicate the student is proficient in the subject every year.
(d) For every student to answer enough questions to ensure the student is on track to being proficient in the subject by a certain year (e.g., two or three years in the future).

1C. Please rate how confident you are that your response to the question above is correct.

2. For state-mandated tests that students must take in public elementary, middle, and high school, if students do not perform as expected, the Federal government (as opposed to the state or the school) mandates:

(a) The student's teacher move to a grade in which the teacher is better at teaching.
(b) The student gets held back a grade until the student learns enough to pass the test.
(c) The school must purchase new educational materials such as textbooks that are more appropriate for the learning styles of the students at the school.
(d) The school receive a penalty, such as being required to provide tutoring to all students, firing administrators at the school, or closing the school altogether.

2C. Please rate how confident you are that your response to the question above is correct.

Not confident			Moderately confident		Completely confident	
1	2	3	4	5	6	7

3. The state-mandated tests that students must take in public elementary, middle, and high school are created to align to what students are supposed to learn according to:

(a) The student's teacher.
(b) The student's school.
(c) The state in which the student's school is located.
(d) The U.S. Department of Education.

3C. Please rate how confident you are that your response to the question above is correct.

Not confident			Moderately confident		Completely confident	
1	2	3	4	5	6	7

4. Which of the following most accurately describes the goal of the No Child Left Behind Act, which is the Federal law that mandates state tests that students must take in public elementary, middle, and high school?

(a) The act is specifically designed to help ensure the United States has a more competitive science and technology workforce compared to emerging nations such as China and India.
(b) The act is specifically designed to ensure that all students in the United States are meeting the same national standards of learning in academic areas including math, reading, and science.
(c)The act is specifically designed to ensure no student is left without the critical resources that are needed to learn, including current textbooks, laboratory equipment for science classes, and at least some access to the Internet within the school building.
(d) The act is specifically designed to ensure that all students have access to an adequate education as defined by each individual state.

4C. Please rate how confident you are that your response to the question above is correct.

Not confident			Moderately confident		Completely confident	
1	2	3	4	5	6	7

5. For state-mandated tests that students must take in public elementary, middle, and high school, there are certain levels of proficiency. If a student scores at the "proficient" level or higher, the student is said to:

(a) Have enough knowledge and skill to be successful in the next grade level.
(b) Be on track to not take remedial courses in college.
(c) Be sufficiently proficient to succeed in college.
(d) Have mastered grade level work as defined by the state.

5C. Please rate how confident you are that your response to the question above is correct.

Not confident			Moderately confident			Completely confident
1	2	3	4	5	6	7

6. For state-mandated tests that students must take in public elementary, middle, and high school, the Federal government requires the following be publicly available via a school's "report card", which must be provided to parents and is often featured on the Internet:

(a) The average scores for all teachers in a school, separated out by subject area and whether or not the teacher is new to the teaching profession.
(b) The individual scores for all students in the school, although the names of individual students are kept private.
(c) The individual scores for those students in a school whose scores were not considered "proficient", although the names of individual students are kept private.
(d) The average score across all students in each grade in a school, as well as the average score across all students in each of four subgroups race/ethnic subgroups in a grade (African American, Asian / Pacific Islander, Caucasian, Hispanic / Latino).

6C. Please rate how confident you are that your response to the question above is correct.

Not confident			Moderately confident			Completely confident
1	2	3	4	5	6	7

7. For state-mandated tests that students must take in public elementary, middle, and high school, the Federal government mandates that students' scores are used in conjunction with the following when evaluating the effectiveness of a school:

(a) The financial resources available to the school, especially state and local budget allocations.
(b) The socio-economic status of students at the school, especially the level of parents' education and parental involvement in the school.
(c) School characteristics, especially class size and the location of the school in relation to urban or rural areas.
(d) Test scores are the only factors used to evaluate the effectiveness of schools.

7C. Please rate how confident you are that your response to the question above is correct.

Not confident			Moderately confident			Completely confident
1	2	3	4	5	6	7

8. For state-mandated tests that students must take in public elementary, middle, and high school, on average the amount of time that students spend taking the actual state test (i.e., excluding practice tests) is:

(a) 1.0% of the school year
(b) 7% of the school year
(c) 12% of the school year
(d) 18% of the school year

8C. Please rate how confident you are that your response to the question above is correct.

Not confident			Moderately confident			Completely confident
1	2	3	4	5	6	7

9. For state-mandated tests that students must take in public elementary, middle, and high school, the tests are designed to measure:

(a) What the teacher expects the student to learn.
(b) What the school /school district expects the student to learn.
(c) What the state expects the student to learn.
(d) What the Federal government expects the student to learn.

9C. Please rate how confident you are that your response to the question above is correct.

Not confident			Moderately confident			Completely confident
1	2	3	4	5	6	7

Abstract

This mixed method study explored the professional competencies that administrators expect from entry-, mid-, and senior-level professionals as reflected in 1,759 job openings posted in 2008. Knowledge, skill, and dispositional competencies were identified during the qualitative phase of the study. Statistical analysis of the prevalence of competencies revealed significant differences between major functional areas and requirements for educational and work experience. Implications for institutional leaders, graduate faculty, and professional development planning as well as for mixed methods research are discussed.

AUTHORS

John L. Hoffman, Ph.D.
California State University, Fullerton

Marilee J. Bresciani, Ph.D.
San Diego State University

CORRESPONDENCE

Email

jhoffman@fullerton.edu

AUTHORS' NOTE

The authors wish to thank The Placement Exchange and its supporting professional associations for providing the data for this study. The authors additionally thank Chieh-Hsing Chou and Michelle Cuellar for their assistance with coding the data.

Identifying What Student Affairs Professionals Value: A Mixed Methods Analysis of Professional Competencies Listed in Job Descriptions

"When situations change on a dime, as is frequently the case in today's economy, people are what make a difference" (Whitely, as cited in Grund, 2009, p. 12). The student affairs profession values people, and in an era of assessment and accountability, it must also be a profession that values the development and demonstration of competence by those people. Palomba and Banta (1999) defined assessment as "the systematic collection, review, and use of information about education *for the purpose* of [emphasis added] improving student learning and development" (p. 4). Closing the assessment loop from data collection and analysis to improving learning and development often entails changes in the design and delivery of educational programs, but this effort should also consider the knowledge, skills, and dispositions of the educators who enact those programs. For this reason, assessment-minded scholar-practitioners have recently afforded increased attention to the question of what specific competencies successful student affairs professionals need. The research and professional literature addressing this question has largely emphasized new professional competencies and relied on self-reports from senior student affairs officers, mid- and entry-level professionals, and graduate preparation faculty. Though Pace (1985) suggested that self-report data in well-designed studies are valid, Brener, Billy, and Grady (2003) have shown that cognitive and situational factors may influence participant self-report responses. Thus, it is important to triangulate self-report data with additional measures of the phenomenon in question.

The purpose of this exploratory mixed methods study is to extend current literature addressing the self-reported competencies required of entry-, mid-, and senior-level student affairs professionals by examining which competencies college and university administrators include in formal job postings. While we acknowledge that the content of

job postings is often politically negotiated between competing campus interests, influenced by campus culture, or tempered by human resource professionals who wish to avoid potential litigation, we also assert that job postings are a meaningful reflection of the competencies college and university administrators desire from the professionals they hire. Further, we stress both the functional and the symbolic importance of the content of job postings. Functionally, the content of job postings provide an initial indicator to prospective candidates of the outcomes that institutions will expect of them as well as of the competencies that administrators believe will be necessary to achieve those outcomes. Symbolically, job postings present a first impression of the underlying institutional values that will guide the desired means of accomplishing outcomes. Thus, the findings of this study should inform not only the work of the preparation programs and professional development organizations that help student affairs professionals to develop the competencies necessary to be successful, it should also inform the work of the leaders who design job postings and their associated job descriptions.

To address the study's purpose, the researchers posed these research questions: (a) Which competencies do college and university administrators most frequently include in formally advertised job postings? (b) Which competencies are more or less often required of student affairs professionals *within* various functional areas? (c) What differences exist in competency requirements *between* major functional areas, different types and sizes of institutions, and positions requiring different levels of education and work experience?

Literature Review

Our framework for this study considered both the scope and development of competencies through education and professional experience. Standards developed by the American Psychological Association (APA), the Council for Accreditation of Counseling and Related Educational Programs (CACREP), and the National Association of Social Workers (NASW) emphasize *knowledge and skills*. Because the scope of student affairs work includes leadership and educational functions as well as advocacy and helping roles, we chose to also consider *dispositional* competencies. The literature addressing dispositions is grounded in works such as Goleman's (1995) model of emotional intelligence and Perkin's (1993) work addressing the interconnections between neurological, experiential, and reflective intelligence and spans both education and leadership studies (e.g. Avolio, 2010; Bass & Riggio, 2005; Leithwood, Seashore Louis, Anderson, & Wahlstrom, 2004). For the purpose of this study, we defined dispositional competencies as encompassing "attitudes, values, and beliefs" (NCATE, 2008, p. 80) and "habits of the mind…that filter one's knowledge, skills, and beliefs and impact the action one takes in professional settings" (Thornton, 2005, p. 62).

Symbolically, job postings present a first impression of the underlying institutional values that will guide the desired means of accomplishing outcomes.

To address the evaluation and development of competencies, we drew upon the five stages of the skill acquisition model developed by Dreyfus and Dreyfus (1980). Movement through the various stages is marked first by a shift from following concrete rules (stage one, *novice*) to identifying recurrent patterns or aspects (*competence)*. In the third stage, *proficiency,* professionals move beyond aspect recognition to evaluating aspects in the context of various situations. Professionals next demonstrate *expertise* when intuition replaces aspect recognition and evaluation. Finally, the fifth stage of *mastery* is marked by a degree of transcendence of expertise.

Identifying Competencies

American College Personnel Association (ACPA) and National Association of Student Personnel Administrators (NASPA; 2010) recently published a set of 10 professional competency areas for student affairs professionals. Their work built on prior sets of competencies identified by the Counsel for the Advancement of Standards in Higher Education (CAS, 2006) and ACPA (2008), as well as numerous empirical studies. This work, officially

adopted by ACPA and NASPA governing boards, extends prior discussions of professional competence by providing outcomes and descriptions of each competency area that were "divided into *basic, intermediate,* and *advanced levels* that delineate the increasing complexity and ability that should be demonstrated by practitioners as they grow in their professional development" (ACPA & NASPA, 2010, p. 4). The 10 competency areas are (a) advising and helping, (b) assessment, evaluation, and research, (c) equity, diversity, and inclusion, (d) ethical professional practice, (e) history, philosophy, and values, (f) human and organizational resources, (g) law, policy, and governance, (h) leadership, (i) personal foundations, and (j) student learning and development.

Current competency research is largely limited to expectations for new professionals, but not those for mid- or senior-level professionals who should be able to demonstrate skill acquisition at a more advanced level.

Several recent research studies have also aimed to identify sets of professional competencies that entry-level professionals should possess. Lovell and Kosten (2000) conducted a meta-analysis of 30 years of research in order to identify 16 broad knowledge, skill, and personal trait characteristics that were vital to success in the student affairs profession. Their competencies were similar to those recently generated by ACPA and NASPA, though they did not include competencies in the areas of equity, diversity, and inclusion; ethical professional practice; or history, philosophy, and values. Burkard, Cole, Ott, and Stoflet (2005) employed a Dephi design involving multiple iterations of surveys with a panel of 104 mid- and senior-level student affairs administrators. The 32 competencies identified by Burkard et al. aligned well with the ACPA and NASPA competencies, though they did not include any competencies in the areas of ethical professional practice or history, philosophy, and values. These competency areas did materialize in a recent study by Hickmott and Bresciani (in press) who classified the 26 competencies that emerged from their study as knowledge, skills, or dispositions. In this study, ethical practice was included with the legal knowledge competency. The Hickmott and Bresciani study differed from those conducted by Lovell and Kosten and Burkard et al. in that they employed a grounded theory approach to analyze formal documents from 54 graduate preparation programs. Thus, competencies related to ethical professional practice and history, philosophy, and values emerged from the study that examined graduate program documents but not from those that examined prior research (Lovell & Kosten, 2000) or the self-reports of mid- and senior-level practitioners (Burkard et al., 2005).

Development and Evaluations of Competencies

Additional research has endeavored to assess faculty and administrator perceptions regarding the degree to which entry-level professionals have mastered essential competencies. Herdlein (2004) administered a mixed methods survey to a sample of 48 senior student affairs officers (SSAOs) who worked at colleges and universities with student affairs graduate preparation programs. Herdlein found that SSAOs were generally satisfied with the level of new professional competence, and that they rated new professionals highest in the areas of overall knowledge of higher education, knowledge of student development theory, and skills in leadership, technology, and counseling. These SSAOs rated new professionals lowest in the skill areas of budgeting, strategic planning, and research and assessment, as well as in the knowledge areas of campus politics and legal matters. More recently, Herdlein, Kline, Boquard, and Haddad (2010) studied faculty perceptions of the importance of various learning outcomes for their programs. When responding to survey items, the most highly rated learning outcomes were in the areas of (a) knowledge of student characteristics and the effects of college, (b) student development theory, (c) how values inform practice, and (d) multicultural perspectives. The lowest rated outcomes were for (a) teaching methods, (b) international education, (c) governance and public policy, and (d) research methods. When asked via an open-ended question to identify the course that was most important to professional practice, faculty listed student development and learning courses more than twice as often as any other course.

Waple (2006) studied entry-level professionals themselves rather than SSAOs or faculty. Waple's findings largely mirrored Herdlein's (2004), though Waple found that new professionals rated themselves lower in several technology-related competency areas. Cuyjet, Longwell-Grice, and Molina (2009) studied recent preparation program graduates and their

supervisors and found that graduates rated their knowledge acquisition higher than did their supervisors. Renn and Jessup-Anger (2008), however, found that assessing their own levels of competence and proving themselves were significant challenges for new professionals.

Two additional studies sought to explore differences in the perceptions of SSAOs and faculty regarding entry-level competencies. Kuk, Cobb, and Forrest (2007) analyzed survey responses from 60 SSAOs, 60 mid-level managers, and 60 faculty regarding the importance of 50 competencies that aligned with four broad clusters of knowledge and skill competencies. They found that faculty rated the importance of (a) individual practices and administration, (b) goal setting and the ability to manage change, and (c) managing organizations and groups significantly lower than did either SSAOs or mid-level managers; they found no differences for professional knowledge and content. Faculty were also more likely to expect entry-level professionals to master professional knowledge and content through coursework, though they expected them to learn how to manage organizations and groups in professional settings. Renn and Jessup-Anger (2008), in their grounded theory study of the experiences of entry-level professionals, found that new professionals desired greater support in managing the cultural dynamics of work environments.

Dickerson et al. (2011) compared ratings by 125 faculty and 275 SSAOs of 51 discrete knowledge, skill, and dispositional competencies. They found no differences between SSAOs and faculty in the perceived importance of 49 competencies and no differences in assessments of the degree to which new professionals possessed 42 of the 51 competencies. Dickerson et al. further examined differences between the degree to which the entire sample rated the competencies as "desired for" and "currently possessed by" new professionals. They found significant gaps in the areas of fiscal management, assessment, and knowledge of legal standards, findings that mirror those of Herdlein (2004) and Waple (2006). However, Dickerson et al. also found significant gaps for collaboration, conflict management, the application of theory to practice, and written communication, areas identified as strengths among the Herdlein and Waple studies.

To summarize, there appears to be emerging consensus within current research and professional literature regarding the scope of knowledge, skill, and dispositional competencies for entry-level professionals. However, this consensus largely reflects analyses of the self-reports by SSAOs, faculty, and other practitioners regarding these competencies, but not which competencies administrators include in job postings. Further, current competency research is largely limited to expectations for new professionals, but not those for mid- or senior-level professionals who should be able to demonstrate skill acquisition at a more advanced level.

Methodology

We drew from a pragmatic orientation to adapt what Creswell and Plano Clark (2011) described as an exploratory mixed methods research design. This design involves an initial qualitative data collection and analysis phase that informs subsequent quantitative data collection and analyses. In keeping with what Patton (1990) identified as a "mixed form design" and Tashakkori and Teddlie (1998) described as a "mixed model design," our study involved a single data set that we initially analyzed using a qualitative approach. The results of the initial qualitative analysis were then analyzed using quantitative methods. This mixing of data analyses allowed us to extend the identification of competencies from the data set to an exploration of the prevalence of these competencies both within and between groups inside the larger data set. The following provides an overview of the data utilized in this study. Because of the sequential nature of the study, we present the research design and results for each phase of the study separately.

Data Collection and Sample

The data for this study consisted of all 1,759 job descriptions posted through *The Placement Exchange* (TPE) in 2008. TPE is a partnership between NASPA, the Association of College and University Housing Officers-International (ACUHO-I), the National Association for Campus Activities (NACA), the Association for Student Judicial Affairs (ASJA), the

Table 1

Institution and Job Characteristics

Category	*N*	%
Institution Type (n = 1,759)		
2-Year Institutions	116	6.6%
4-Year Public Institution	600	34.1%
4-Year Private Institution	925	52.6%
Professional Organization/Other	118	6.7%
Institution Size (n = 1,759)		
Under 5,000	658	37.4%
5,000-9,999	287	16.3%
10,000-20,000	372	21.1%
More than 20,000	442	25.1%
Job Category (n = 1,729)*		
Admissions and Enrollment Management	45	2.6%
Career Services	70	4.0%
Counseling, Health, and Wellness	65	3.7%
Greek Life	64	3.6%
Judicial Affairs	50	2.8%
Multicultural (including lesbian, gay, bisexual/transgendered services, and women's centers)	105	6.0%
Orientation and New Student Programs	63	3.6%
Residence Life	797	45.3%
Student Activities (including leadership programs and student unions)	232	13.2%
Student Affairs Administration	405	23.0%
Other Student Affairs Positions	150	8.5%
Highest Level Education Preferred or Required (n = 1,668)**		
Bachelor's Only	129	7.8%
Master's Preferred	611	37.1%
Master's Required	700	42.5%
Doctorate Preferred or Required	208	12.6%
Years of Professional Experience Required (n = 1,422)		
0-3 Years	779	54.8%
4-6 Years	512	36.0%
7 or More Years	131	9.2%

Note.

*219 job postings were categorized under two or functional areas; thus, the cumulative percentages for this section exceed 100%.

**The data for some job postings did not include highest level of education preferred or years of professional experience required.

National Orientation Directors Association (NODA), the Association of Fraternity/Sorority Advisors (AFA), and HigherEdJobs.com (www.theplacementexhange.org). TPE holds an annual placement conference prior to the national meeting of NASPA and serves as a centralized online web source for student affairs job postings. Data collected for this set of job postings included the institutional type and size, job category, education and work experience required of applicants, and the full text of the job postings (See Table 1).

Assumptions and Limitations

This single data set served as our source and therefore we assumed that the sample was representative of student affairs positions throughout the United States. In qualitative terms, these job postings served as a large data set that should contribute to reasonable external or "ecological" validity (see Gall, Borg, & Gall, 1996), meaning that we anticipated that a similar set of competencies would emerge from a grounded theory analysis of an alternate comprehensive set of student affairs job postings. Since there are no existing data that accurately break down the number of student affairs professionals employed nationally at various types and sizes of institutions or in various functional areas, it was impossible to compare this sample to the full population of student affairs jobs.

There are several limitations to our assumptions regarding the representativeness of the sample and to the external validity of findings. The data were collected in 2008, just prior to a significant economic downturn and shortly following the publication of the Spelling's Report (U.S. Department of Education, 2006). This, along with other socio-historical factors likely influenced the content of some job postings; one should use some caution, for example, in assuming that job postings advertised during the economic downturn would reflect the findings of this study. Further, a visual review of Table 1 reveals that positions within community colleges were underrepresented, as were positions within the functional areas of admissions and enrollment, academic advising, outreach, and financial aid. Additionally, there was significant variation in the content, length, and detail of information included in job postings and descriptions. We assumed that these variations, which are a form of measurement error, were randomly distributed across the large sample of data.

Identifying Competencies

For the initial phase of the study, we employed an adaptation of open and substantive coding to identify categories of competencies that were emergent within the data set (see Morse, 2009). Because our first research question aimed to identify competencies but not to explore the interrelationships between them, we utilized only open and substantive coding processes. We delimited the open coding process to the first 100 job postings in the data set and used these data as the basis for identifying competency areas.
Twenty-three job competency categories emerged from our initial analysis and clustering of codes. Drawing from the job postings associated with each category, we generated definitions for each competency area and then used these definitions to re-code the entire data set of 1,759 job postings. For this final re-coding, we used whole job postings as the unit of analysis; in effect, we assigned a yes or no dummy code for each of the 23 competency areas to every job posting. We then reviewed frequencies to ensure discriminant validity between the various competency areas. As there was more than 90% overlap among the job postings coded as "assessment," "program evaluation," and "research," we collapsed these three into a single competency category. Table 2 summarizes the emergent definitions and frequency counts for each of the final 21 competencies.

Knowledge of the profession's history and philosophy emerged as a stand-alone competency in this study, a finding that suggests that this competency is important to more than just faculty.

Table 2

Emergent Competency Areas

Competency	Definition	Frequency	Percentage
Programming	This consisted of requirements to design, develop, implement, or manage programs, activities, and other forms of learning interventions.	1,139	64.8%
Communication	This encompassed both written and oral forms of communication as well as interpersonal communication.	1,038	59.0%
Assessment, Evaluation, and Research	This area included all references to outcomes-based assessment, program review and evaluation, research, and data analysis.	846	48.1%
Teaching and Training	This involved teaching credit-bearing and non-credit courses as well as providing various forms of training.	756	43.0%
Leadership	This encompassed efforts to provide long-term direction to divisions or departments including planning, forecasting, and visioning.	753	42.8%
Budgeting and Fiscal Management	This included all references to understanding and managing budgets, financial plans, and fiscal resources. It did not include fundraising.	736	41.8%
Collaboration with Non-Faculty Professionals	This involved requirements for teamwork, partnership, or collaboration that specifically cited non-faculty professionals on the campus. (Collaboration with faculty was also cited in 480, 69.3%, of these postings.)	693	39.4%
Collaboration with Faculty	This included requirements for teamwork, partnership, or collaboration specifically with faculty. (Collaboration with non-faculty was also cited in 479, 73.8%, of these postings.)	649	36.9%
Law and Policy	This consisted of knowledge of higher education law, oversight of disciplinary procedures, development and administration of policy, and understanding governance structures.	615	35.0%
Diversity and Social Justice	This entailed required knowledge, skill, or dispositional competencies necessary to work with diverse students or to create inclusive learning environments. This extended beyond mere descriptions of working in a diverse environment or identification of the institution as an equal opportunity employer.	600	34.1%
Technology	This consisted of all requirements to use computers and other forms of technology.	589	33.5%
Student Learning and Development	This included required knowledge of student development or learning theory and skill-based expectations to apply these in practice.	552	31.4%
Practical Competence	This consisted of an array of skills related to time management, organizational skills, managing multiple responsibilities, working autonomously, and meeting deadlines.	423	24.0%
Advising Groups	This encompassed requirement for any form of advising student clubs, organizations, or groups of students.	318	18.1%
Crisis Management	This included all requirements to respond to emergencies, assess risks, or manage crisis situations.	318	18.1%
Attitudes and Dispositions	This covered broad requirements for dispositional competencies; the most common were creativity, enthusiasm, flexibility, and a positive attitude.	252	14.3%
Advising Individuals	This encompassed requirements for providing any form of advising or counseling to individual students.	210	11.9%
Conflict Mediation	This included requirements for mediating conflicts between individuals or groups of students.	174	9.9%
Foundations of the Profession	This entailed all requirements for knowledge of the history, philosophy, or ethical standards of student affairs or higher education.	100	5.7%
Fundraising	This included requirements to engage in grant-writing, organize fundraising events, and solicit donations.	60	3.4%
Critical Thinking	This encompassed requirements for problem-solving, decision-making, critical thinking, and reflective practice.	56	3.2%

Testing for Differences

For the second phase of the study, we constructed a series of cross-tabulation (crosstab) tables to compare the frequencies *within* various functional areas as well as *between* functional areas, institutional types and sizes, and sets of required education and work experience. In what follows, we review the specifics of our research design followed by the results for each research question sequentially.

Differences within Functional Areas

To examine differences within the various functional areas, we first delimited the sample to the 1,540 job postings categorized in only one functional area. Given 21 competency areas (each coded yes or no) and 11 functional areas (coded yes or no), this yielded 231 2x2 crosstab tables. We used Fisher's exact test to check for differences and Phi to test for effect size. The null hypothesis for Fisher's exact assumes that the prevalence of yes and no values for each competency area will be divided proportionally across each of the 11 functional areas. When statistically significant, we rejected the null hypothesis and assumed that there were differences in the prevalence of the competency for the given functional area. In those instances, we further calculated Phi to examine the effect size of the differences. Phi is a symmetric measure that determines the effect size of differences in 2x2 crosstab tables. Cohen (1988) placed Phi (along with the other symmetric measures used in this study) into the same family of statistics as the more common Pearson's *r* measure of correlation; thus, one should interpret the Phi statistic in a similar manner as one would interpret a Pearson's *r*.

Table 3 reviews the results of tests of differences within each of the functional areas. Because non-parametric measures are sensitive to sample size, one should not compare Phi values *between* two different functional areas. The Phi values are an accurate measure of the effect size for differences *within* each of the functional areas. While there were many statistically significant differences in competency prevalence within the various functional areas, the effect size of these differences were generally small or quite small. According to Cohen (1988), effect sizes for Phi that are less than .10 are much smaller than typical for the social sciences; those between .10 and .30 are small.

Differences between Groups

Our third research question addressed differences between various functional areas, between different types and sizes of institutions, and between positions requiring different levels of education and work experience. To compare the effect size for differences between various functional areas using non-parametric statistics, one needs to have sufficient numbers within each of the comparison groups, and there should not be any large fluctuation in the sample sizes of the groups. Given the large sample size differences for each of the functional areas, it was impossible to run comparative data without heavily weighting the data, which would significantly increase the likelihood of Type 1 measurement error. For this reason, we limited comparative analyses to institutional type, institutional size, and the levels of education and work experience required.

Differences by institutional type. In order to test for differences between institutional types, we first delimited our sample to 2-year, 4-year private, and 4-year public institutions (n=1,641) and then weighted the data for the purpose of comparison. The result was 21 2x3 tables, one for each competency; each table analyzed the competency (yes or no) against the three institution types (2-year, 4-year private, or 4-year public). For 2x3 tables with nominal data, the chi-square is the appropriate non-parametric test of difference and Cramer's V is the preferred symmetric measure.

A few statistically significant differences emerged in the comparisons by institutional type. Two-year institutions were more likely to include requirements for collaboration with other professionals, $\chi 2(2) = 49.36$, $p < .001$; $V = .13$, $p < .001$; but less likely to include competencies related to crisis management, $\chi 2(2) = 57.67$, $p < .001$; $V = .15$, $p <$

Table 3

Differences Within Functional Areas

Functional Area/Competency	Fisher's Exact Test Sig.	Phi Value	Phi Sig.
Admissions/Enrollment Management (n = 30)			
Technology	.00	.08	.00
Advising Groups	.00	-.07	.01
Crisis Management	.00	-.07	.01
Student Learning and Development	.00	-.09	.00
Teaching and Training	.00	-.12	.00
Career Services (n = 55)			
Advising Individuals	.00	.16	.00
Technology	.00	.14	.00
Collaboration with Faculty	.00	.08	.00
Communication	.00	.07	.00
Crisis Management	.00	-.09	.00
Law and Policy	.00	-.12	.00
Counseling/Health/Wellness (n = 40)			
Advising Groups	.00	-.07	.01
Greek Life (n = 47)			
Advising Groups	.00	.18	.00
Judicial Affairs (n = 26)			
Conflict Mediation	.00	.09	.00
Law and Policy	.00	.09	.00
Teaching and Training	.00	.09	.00
Assessment, Evaluation, and Research	.01	.07	.01
Budgeting and Fiscal Management	.01	-.07	.01
Multicultural Services (n = 63)			
Diversity and Social Justice	.00	.12	.00
Fundraising	.00	.11	.00
Crisis Management	.00	-.08	.00
Law and Policy	.00	-.09	.00
Residence Life/Housing (n = 720)			
Crisis Management	.00	.24	.00
Conflict Mediation	.00	.14	.00
Law and Policy	.00	.09	.00
Advising Groups	.00	.08	.00
Foundations of the Profession	.01	-.07	.00
Programming	.00	-.08	.00
Leadership	.00	-.10	.00
Teaching and Training	.00	.13	.00
Collaboration with Faculty	.00	-.11	.00
Communication	.00	-.13	.00
Technology	.00	-.13	.00
Collaboration with Other Professionals	.00	-.14	.00
Budgeting and Fiscal Management	.00	-.15	.00
Assessment, Evaluation, and Research	.00	-.16	.00
Fundraising	.00	-.16	.00
Student Activities (n = 158)			
Advising Groups	.00	.15	.00
Budgeting and Fiscal Management	.00	.09	.00
Fundraising	.01	.08	.00
Communication	.01	-.07	.00
Conflict Mediation	.00	-.09	.00
Diversity and Social Justice	.00	-.10	.00
Crisis Management	.00	-.11	.00
Collaboration with Faculty	.00	-.12	.00
Collaboration with Other Professionals	.00	-.14	.00
Student Affairs Administration (n = 287)			
Leadership	.00	.18	.00
Budgeting and Fiscal Management	.00	.16	.00
Collaboration with Faculty	.00	.16	.00
Collaboration with Other Professionals	.00	.15	.00
Diversity and Social Justice	.00	.11	.00
Communication	.00	.10	.00
Foundations of the Profession	.00	.09	.00
Fundraising	.01	.08	.00
Assessment, Evaluation, and Research	.01	.07	.01
Advising Individuals	.01	-.07	.01
Advising Groups	.00	-.16	.00
Teaching and Training	.00	-.17	.00
Other Student Affairs Positions (n = 77)			
Fundraising	.00	.12	.00
Technology	.00	.10	.00
Communication	.01	.07	.01
Advising Groups	.00	-.07	.01
Crisis Management	.00	-.08	.00

Note. Table 3 lists competency areas that were statistically significant at $p < .01$ listed from most common to least common according to Phi values (positive Phi values indicate that the competency was more likely to occur within the functional area; negative Phi values indicate the opposite)

.001; teaching and training, $\chi2(2) = 53.69$, $p < .001$; $V = .14$, $p < .001$; group advising, $\chi2(2) = 89.54$, $p < .001$; $V = .18$, $p < .001$; and individual advising, $\chi2(2) = 55.34$, $p < .001$; $V = .14$, $p < .001$. Private 4-year institutions were more likely than public 4-year institutions to include requirements within these latter two advising competency areas, as well as in conflict mediation, $\chi2(2) = 30.86$, $p < .001$; $V = .11$, $p < .001$.

Differences by institutional size. For institutional size, we did not weight the data because each group had at least 200 cases and there were minimal sample size differences between the groups. The result was 21 2x4 tables, one for each competency; each table analyzed the competency (yes or no) against the four institution size groups (less than 5,000; 5,000-9,999; 10,000-20,000; or more than 20,000). Given 2x4 crosstab tables with ordinal data, we used the chi-square to test for differences and Kendall's tau-b to examine effect sizes. Few statistically significant results emerged from these analyses. The most significant difference was for the attitudes and dispositions competency, which was slightly more prevalent at smaller institutions, $\chi2(3) = 17.58$, $p < .01$; tau-b = -.09, $p < .001$.

We encourage leaders who design job postings and their associated job descriptions to more intentionally and systematically include assessment-related competencies in these important documents.

Differences by level of education. Though we were unable to classify the various job postings as entry-level, mid-level, or senior-level, the requirements for education and work experience afforded us a proxy to examine differences along a range of positions extending from entry-level to senior-level. For education requirements, we organized the 1,648 positions that included education requirements into 21 2x4 crosstab tables, one for each competency; each table analyzed the competency (yes or no) against the four levels of required education (bachelor's only, master's preferred, master's required, or doctorate preferred/required). We used the chi-square to test for differences and Kendall's tau-b to test the effect size. Table 4 presents the results of these analyses. Not all tau-b values were statistically significant; this reflects instances where there were differences among the four groups that do not reflect the ordinal progression of the four educational levels (e.g. when a competency was more prevalent among the master's required and preferred groups than either the bachelor's only or the doctorate preferred/required group). We listed the competency areas in Table 4 in rank order from those positions requiring the most education to those requiring the least.

Table 4

Differences Between Positions Requiring Different Levels of Education

Competency	Chi-Square Value (df=3)	Chi-Square Significance	Kendall's tau-b Value	Kendall's tau-b Significance
Leadership	202.7	.00	.22	.00
Budgeting and Fiscal Management	172.7	.00	.20	.00
Collaboration-Faculty	100.8	.00	.15	.00
Assessment, Evaluation, and Research	74.3	.00	.14	.00
Diversity and Social Justice	47.1	.00	.09	.00
Collaboration with Other Professionals	46.9	.00	.09	.00
Student Learning and Development	28.6	.00	.08	.00
Law and Policy	33.8	.00	.07	.00
Communication	20.1	.00	.06	.00
Crisis Management	41.3	.00	.05	.00
Foundations of the Profession	33.6	.00	.05	.01
Fundraising	32.2	.00	.05	.02
Conflict Mediation	11.9	.01	.03	.06
Teaching and Training	39.8	.00	-.01	.53
Advising Groups	77.4	.00	-.06	.00
Advising Individuals	24.1	.00	-.07	.00
Practical Competence	39.0	.00	-.10	.00
Technology	50.5	.00	-.11	.00

Note. Positive Kendall's tau-b values indicate that the competency area was more prevalent for positions requiring higher levels of education; negative values indicate the opposite.

Differences by level of work experience. We organized the 1,422 postings that included work experience requirements into 21 2x3 tables, one for each competency; each table compared the competency (yes or no) against the three levels of work experience required (0-3 years, 4-6 years, or 7 or more years). We used the chi-square to test for differences and Kendall's tau-b to test the effect size. Table 5 presents the results. We listed the competency areas in Table 5 in rank order from those positions requiring the most work experience to those requiring the least.

Discussion

Identifying Competencies

The 21 competencies that emerged from this study aligned well with those generated by ACPA and NASPA (2010), as well as with those identified by the empirical studies that informed the ACPA/NASPA publication (e.g. Burkard et al., 2005; Cuyjet et al., 2009; Herdlein, 2004). Interestingly, one could align three of the four most prevalent competencies in this study (programming, communication, and teaching and training) with ACPA/NASPA's "advising and helping" competency. As noted previously in the literature review, ethical practice – one of the ACPA and NASPA competencies – emerged only from the Hickmott and Bresciani (in press) study of graduate preparation curricula, and then only when

Table 5

Differences Between Positions Requiring Different Levels of Work Experience

Competency	Chi-Square Value (df=2)	Chi-Square Significance	Kendall's tau-b Value	Kendall's tau-b Significance
Leadership	240.3	.00	.30	.00
Budgeting and Fiscal Management	190.6	.00	.27	.00
Collaboration with Other Professionals	107.0	.00	.20	.00
Collaboration with Faculty	79.6	.00	.17	.00
Diversity and Social Justice	49.3	.00	.13	.00
Assessment, Evaluation, and Research	44.4	.00	.12	.00
Communication	34.2	.00	.11	.00
Fundraising	21.9	.00	.09	.00
Programming	15.9	.00	.08	.00
Foundations of the Profession	15.2	.00	.07	.00
Law and Policy	23.4	.00	.07	.00
Technology	8.4	.02	-.01	.74
Crisis Management	15.9	.00	-.07	.00
Conflict Mediation	17.7	.00	-.08	.00
Advising Individuals	44.9	.00	-.12	.00
Teaching and Training	109.6	.00	-.19	.00
Advising Groups	151.8	.00	-.24	.00

Note. A positive Kendall's tau-b value indicates that the competency area was more prevalent for positions requiring higher work experience; a negative value indicates the opposite.

integrated with legal knowledge. It did not emerge from either the Lovell and Kosten (2000) or the Burkard et al. (2005) study, and it did not emerge as a stand-alone competency in this study. This finding highlights the fact that the findings of this study were descriptive, not prescriptive in nature. The fact that ethical practice did not emerge as an important competency area does not mean that it is not important for professionals or the profession.

ACPA and NASPA's history, philosophy, and values competency was also missing from the Burkard et al. (2005) study of mid- and senior-level perceptions and the Lovell and Kosten (2000) meta-analysis of prior practitioner research, though it did emerge in the Hickmott and Bresciani (in press) study of graduate preparation program documents. One might have interpreted this to mean that history and philosophy are more important to faculty than to practitioners. However, knowledge of the profession's history and philosophy emerged as a stand-alone competency in this study, a finding that suggests that this competency is important to more than just faculty. Further, this competency was most commonly included among positions in student affairs administration that required higher levels of education and experience, which may reflect how knowledge of history and philosophy may contribute to practitioners' capacities in aspect recognition and evaluation, processes associated with higher levels of skill acquisition in the Dreyfus and Dreyfus (1980) model.

Developing and Evaluating Competencies

When comparing the prevalence of these competencies within this sample to studies that have aimed to rank the importance of competencies, several interesting differences emerged. The studies by Burkard et al. (2005), Herdlein (2004), and Waple (2006) each identified sets of attitudes and dispositions, practical skills, and critical thinking skills among the most important or highly ranked competencies in their studies. However, critical thinking (3.2%), attitudes and dispositions (14.3%), and practical competencies (24.0%) were among those least frequently included among the 2008 job postings. This may reflect hesitancy by human resource departments to include in job postings those competencies that are difficult to measure in selection processes or it may mean that these skill sets are assumed. Yet, if these skills are important and serve as criteria for future performance evaluations, administrators may be wise to negotiate for their inclusion in formal job postings. Future research could address this issue by triangulating job posting analyses with performance evaluation criteria or qualitative interviews with the administrators and human resource professionals who craft job postings, job descriptions, and performance evaluation protocols.

On the other end of the spectrum, colleges and universities included assessment, evaluation, and research competencies in 48.1% of 2008 job postings, but the related knowledge and skill competencies were ranked in the middle of the competency sets generated by Waple (2006) and Herdlein et al. (2010) and near the bottom of the 32 competencies generated in the Burkard et al. (2005) study; they were not included at all among the 34 traits identified as critical for success in the Herdlein (2004) study. This may reflect the growing importance of outcomes-based assessment and program review in student affairs and higher education particularly in light of the growing economic challenges and increased calls for accountability. Regardless of the reason, we find the increased prevalence in this study encouraging. We also suggest that competency in assessment, evaluation, and research is both germane to all functional areas, and it should be the work of all professional educators on campus. Thus, we encourage leaders who design job postings and their associated job descriptions to more intentionally and systematically include assessment-related competencies in these important documents.

Functional Area Differences

Many of the competencies most frequently required of student affairs administrators were also common within career services and multicultural services, which suggests that these areas may also serve well as training grounds for senior-level leadership.

Most of the differences within the various functional areas seemed intuitive, though there were a few surprises. As noted previously, the fact that there were statistically significant differences for a greater number of competencies within residence life, student activities, and student affairs administration may reflect a broader set of desired competencies for these positions, but it may also reflect the sensitivity of non-parametric measures to sample size. Among the functional areas with smaller sample sizes, the fact that fundraising emerged as a more prevalent competency within multicultural services was noteworthy. This may reflect the reality that multicultural services is often neither self-supporting, as is the case with residence life, nor supported by student fees, as is often the case with student activities. It may also reflect the growing availability of grants to support those services and programs and the understanding that these programs are desirable philanthropic venues for many donors. However, it may also be that some institutions or divisions of student affairs continue to view the work of multicultural services as more peripheral than central to their mission. In any case, we suggest that future studies explore why higher percentages of positions in multicultural services require fundraising competencies.

In residence life, fundraising competencies were among the least prevalent along with assessment, evaluation, and research competencies, and the two collaboration-related competency areas. This is of interest since residence life has served as a common training ground for advancement in the student affairs profession, yet the competencies in the areas of collaboration, assessment, evaluation, and research, and fundraising were among the more prevalent required of student affairs administrators. Also common for student affairs administration positions were the diversity and social justice competencies, a pattern matched only by the multicultural services functional area. We certainly do not question the value of residence life experience for advancement in the student affairs profession. That said, we note that many of the competencies most frequently required of student affairs administrators were also common within career services and multicultural services, which suggests that these areas may also serve well as training grounds for senior-level leadership.

Competencies for Entry-, Mid-, and Senior-Level Professionals

Several interesting differences emerged between job postings requiring different levels of education and work experience, our proxy for examining differences between entry-, mid-, and senior-level positions. It is important to note that Dreyfus and Dreyfus (1980) described advancements in "skill acquisition" in terms of shifts to more ambiguous, situational, holistic, and intuitive means of functioning. In light of this study's findings, it appears that development of competence in some areas is cumulative, which is an assumption that is consistent with the Dreyfus and Dreyfus model and upheld by the NASPA/ACPA (2010) Professional Competency Area document. This seemed to be the case, for example, with the leadership and fiscal management competencies, which were most frequently included in job postings that required higher levels of education and experience. Other com-

petencies, such as those related to the history and philosophy of the profession, may have greater utility as professionals advance to positions that require more situational, holistic, and intuitive ways of knowing (knowledge), functioning (skill), and being (disposition).

…It appears that the development of competence in some areas is cumulative...

Competency areas such as technology, practical competence, and advising were more prevalent among entry-level postings. In the case of the practical and technological competency areas, this may suggest that these are "gateway" competencies or it could assume that these are competencies that SSAOs delegate to their staff. Practitioners who do not master outcomes for these competencies at what ACPA/NASPA have described as the basic, intermediate, and advanced level may have limited capacity to advance to mid- and senior-level positions. For advising and training competencies, areas that are also more prevalent among entry-level positions, the cumulative level learning may work somewhat differently. One could argue that the skills developed in these competency areas are transferable to the areas of collaboration and leadership. Experience and professional development in advising and training may serve as a precursor to later development within the collaboration and leadership competency areas as well as serving as a prerequisite to career advancement. In light of the attention that ACPA and NASPA (2010) afforded to delineating "the increasing complexity and ability that should be demonstrated by practitioners as they grow in their professional development" (p. 6), we suggest that this progression and development of professional competence is a topic worthy of further investigation.

Implications

Assuming that the scope and prevalence of competencies within advertised job postings reflect the values of administrators in terms of professional education, training, and development, there are important implications for graduate preparation programs and professional organizations, as well as for employers. We invite readers to question whether this listing would accurately represent what the profession values as a whole, albeit in 2008 or beyond. In addition, we encourage practitioners to question where and when practitioners should master these competencies.

Graduate preparation programs are important training grounds for new professionals, and they are most effective when informed by quality standards (see Young & Janosik, 2007). The results of this study, along with the findings of related studies, should inform graduate preparation faculty of the competencies that are most relevant to entry-level and mid-level professionals. For example, the importance of assessment, evaluation, and research and of student learning and development were each evident in this study. Bresciani (2010) has found that training for outcomes-based assessment is most effective when paired with training in student development theory. The findings here further suggest that the competency areas of individual and group advising, conflict mediation and teaching and training may be of particular importance for master's level programs. Future studies should build on the work of Bresciani and the findings of this study to explore synergistic opportunities in training for multiple competency areas.

The integration of training for multiple competency areas is likely even more important for doctoral preparation programs and for those individuals who design and deliver professional development programs for mid-level and senior-level professionals. In these settings, educators should emphasize the development of more ambiguous, situational, holistic, and intuitive competency in the areas of leadership, budgeting and fiscal management, assessment, evaluation, and research, collaboration, and diversity and social justice. Hoffman and Bresciani (2010), for example, found a high co-occurrence of competency requirements in leadership, decision-making, collaboration, and teaching and training for assessment professionals working in student affairs. Paired with the findings of this study, the implication is that best practices for leadership training in student affairs are integrative and sequential. If we assume each competency area to be one that builds upon its expertise over time, then perhaps employers may also want to consider what competencies are required of the profession holistically and how they represent expected competencies in all of their position advertisements and at what level. For guidance in this area, employ-

ers may consider professional literature that is more definitive (i.e., ACPA & NASPA, 2010) than descriptive, as is the case for this study.

We believe that student affairs is a profession that values people, and a profession that values competency within its people. The best of assessment and accountability efforts emphasize systematic self-study for the purpose of improving practices that result in greater levels of student learning and success. Comprehensive assessment efforts that aim to close the loop between self-study and improving practice must consider the knowledge, skill, and dispositional competencies of the educators who design learning interventions for students, both within and outside the classroom. Ongoing research and scholarly discourse regarding the scope and content of competencies will continue to be critical as the student affairs profession intentionally designs and implements professional preparation programs and professional development to educate the people who work so diligently to promote access, equity, and overall student success within higher education.

References

American College Personnel Association. (2008). *Professional competencies: A report of the steering committee on professional competencies.* Retrieved October 1, 2008 from http://www.myacpa.org/au/governance/docs/ACPA_Competencies.pdf

American College Personnel Association & National Association of Student Personnel Administrators. (2010). *ACPA/NASPA professional competency areas for student affairs practitioners.* Washington, DC: Authors.

Avolio, B. J. (2010). *Full range leadership development* (2nd ed.). Thousand Oaks, CA: Sage.

Bass, B. M., & Riggio, R. E. (2005). *Transformational leadership.* Mahwah, NJ: Lawrence Erlbaum Associates.

Brener, N.D., Billy, J.O., & Grady, W.R. (2003). Assessment of factors affecting the validity of self-reported health-risk behavior among adolescents: Evidence from the scientific literature. *Journal of Adolescent Health, 33*, 436-457.

Bresciani, M.J. (2010). Understanding barriers to student affairs professionals' engagement in outcomes-based assess ment of student learning and development. *Journal of Student Affairs, 14*, 81-90.

Burkard, A., Cole, D.C., Ott, M., & Stoflet, T. (2005). Entry-level competencies of new student affairs professionals: A Delphi study. *NASPA Journal, 42*(3), 283-306.

Creswell, J.W., & Plano Clark, V.L. (2011). *Designing and conducting mixed methods research* (2nd ed.). Thousand Oaks, CA: Sage.

Cohen, J. (1988). *Statistical power and analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Council for the Advancement of Standards in Higher Education. (2006). *CAS professional standards for higher education* (6th ed.). Washington, DC: Author.

Cuyjet, M.J., Longwell-Grice, R., & Molina, E. (2009). Perceptions of new student affairs professionals and their supervisors regarding the application of competencies learned in preparation programs. *Journal of College Student Development*, *50*(1), 104-119.

Dickerson, A.M., Hoffman, J. L., Anan, B.P., Brown, K.F., Vong, L.K., Bresciani, M.J., Monzon, R., & Hickmott, J. (2011). A comparison of senior student affairs officers and student affairs preparatory faculty expectations of entry-level professionals' competencies. *The Journal of Student Affairs Research and Practice, 48*(4), 463-479. doi: 10.2202/1949-6605.627

Dreyfus, S. E., & Dreyfus, H. L. (1980). *A five-state model of the mental activities involved in directed skill acquisition.* Unpublished report, University of California, Berkeley.

Gall, M.D., Borg, W.R., & Gall, J.P. (1996). *Educational research: An introduction* (6th ed.). New York, NY: Longman.

Goleman, D. (1995). *Emotional intelligence.* New York, NY: Bantam Books.

Grund, N. (2009). Student affairs and the new economy. *Leadership Exchange*, *7*(3), 11-13.

Herdlein, R. J., III. (2004). Survey of chief student affairs officers regarding relevance of graduate preparation of new professionals. *NASPA Journal, 42*(1), 51-70.

Herdlein, R. J., III, Kline, K., Boquard, B., & Haddad, V. (2010). A survey of faculty perceptions of learning outcomes in master's level programs in higher education and student affairs. *College Student Affairs Journal, 30* (1), 33-45.

Hickmott, J., & Bresciani, M.J. (in press). Examining learning outcomes in student personnel preparation programs. *Journal of College Student Development.*

Hoffman, J.L., & Bresciani, M.J. (2010). Assessment work: Examining the prevalence and nature of learning assessment competencies and skills in student affairs job postings. *The Journal of Student Affairs Research and Practice*, *47*(4), 495-512. doi: 10.2202/1949-6605.6082

Kuk, L., Cobb, B., & Forrest, C. (2007). Perceptions of competencies of entry-level practitioners in student affairs. *NASPA Journal, 44*(4), 664-691.

Leithwood, K. K., Seashore Louis, K., Anderson, S., & Wahlstrom, K. (2004). *How leadership influences student learning.* New York, NY: Wallace Foundation. Retrieved June, 2005, from http://www.wallacefoundation.org/NR/rdon lyres/E3BCCFA5-A88B-45D3-8E27-B973732283C9/0/ReviewofResearchLearningFromLeadership.pdf

Lovell, C.D., & Kosten, L.A. (2000). Skills, knowledge, and personal traits necessary for success as a student affairs administrator: A meta-analysis of thirty years of research. *NASPA Journal, 37*(4), 353-369.

Morse, J. M. (2009). Tussles, tensions, and resolutions. In J. M. Morse, P. N. Stern, & J. Corbin (Eds.), *Developing grounded theory: The second generation* (pp. 13-19). Walnut Creek, CA: Left Coast Press.

National Council for the Accreditation of Teacher Education. (2008, February). *Professional standards for the accreditation of teacher preparation institutions.* Washington, DC: Author. Retrieved from http://www.ncate.org/documents/ standards/ NCATE%20Standards%202008.pdf

Pace, C.R. (1985). *The credibility of student self-reports*. Los Angeles, CA: University of California, The Center for the Study of Evaluation, Graduate School of Education.

Palomba, C.A., & Banta, T.W. (1999). *Assessment essentials: Planning, implementing, and improving assessment in higher education.* San Francisco, CA: Jossey-Bass.

Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd ed.). Newbury Park, CA: Sage.

Perkins, D. (1993). New conceptions of thinking: From ontology to education. *Educational Psychologist, 28*(1), 67-85.

Renn, K.A., & Jessup-Anger, E.R. (2008). Lessons for graduate preparation programs from the National Study of New Professionals in Student Affairs. *Journal of College Student Development, 49*(4), 319-335. doi: 10.1353/csd.0.0022

Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches.* Thousand Oaks, CA: Sage.

Thornton, H. (2006). Dispositions in action: Do dispositions make a difference in practice? *Teacher Education Quarterly, 33*(2), 53-68.

United States Department of Education. (2006). A test of leadership: Charting the future of higher education. Retrieved from http://www.ed.gov/about/bdscomm/list/ hiedfuture/reports/final-report.pdf

Young, D.G., & Janosik, S.M. (2007). Using CAS standards to measure learning outcomes of student affairs preparation programs. *NASPA Journal, 44*(2), 341-366.

Waple, J.N. (2006). An assessment of skills and competencies necessary for entry-level student affairs work. *NASPA Journal, 43*(1), 1-18.

Abstract

Alumni self-ratings of their personal growth were linked to their intellectual development during college four to seven years earlier. Graduates that were satisfied with their personal growth in the arts, creative thinking, making logical inferences, learning independently, exercising initiative, and tolerating other points of view had higher intellectual scores in Commitment and Empathy as undergraduates years earlier. These findings support a relationship between college student intellectual development and alumni perceptions of their personal growth. The implications of this study support continuing the custom of querying graduates about their earlier education, a practice in wide use already; and add to the validity of the Scale of Intellectual Development as a measure of college impact upon personal dispositions.

AUTHOR
T. Dary Erwin, Ph.D.
James Madison University

CORRESPONDENCE

Email
erwintd@jmu.edu

Intellectual College Development Related to Alumni Perceptions of Personal Growth

The role of intellectual development is considered a central component of undergraduate education and receives much attention regarding its nature and importance (Evans, Forney, Guido, Patton, & Renn, 2010; Pascarella & Terenzini, 2005). Many institutions track some elements of intellectual and cognitive development during college, but what effects remain after college is also of fundamental interest. Separate arenas of research activity exist with current students and with alumni, but no longitudinal studies link collegiate intellectual development to personal growth as perceived by alumni. Studying possible connections of collegiate student development with personal growth has implications for programming at the undergraduate level as well as our understanding of the lasting impact of development after college graduation.

Intellectual Development

Intellectual development occupies a key foundation in college impact studies (Collins, 2006; Dawson, 2004; Pascarella, 1985; Torres, 2003; Wang & Rodgers, 2006), and both academic and student affairs programs claim to nurture student's progress in intellectual and cognitive development. A variety of conceptual approaches to cognitive development exist in the literature such as critical thinking, postformal reasoning, and dispositional critical thinking (Evans et al., 2010; Pascarella & Terenzini, 2005). Although critical thinking development is most often defined with cognitive skills such as "identify central issues and assumptions in an argument" (Jones et al., 1995, p. 133), the disposition to think critically has also been noted as part of intellectual development. Some examples of dispositional critical thinking include, "tolerance for new ideas" and "willingness to see complexity in problems" (Facione, Sanchez, Facione, & Gainen, 1995; Taube, 1997).

Recently, the importance of intellectual development has come forth in the Lumina Foundation's project called the Degree Qualifications Profile, which "defines

expected learning outcomes that graduates need for work, citizenship, global participation and life" (Adelman, Ewell, Gaston, & Schneider, 2011, p. 1). Partly motivated by Europe's Bologna Process (2012), this Profile has outlined "broad, integrative knowledge" and "applied learning" skill sets in general education. One illustrative example of intellectual development states, "articulates and defends the significance and implications of his or her specialized work in terms of challenges, trends and developments in a social or global context" (p. 18). This expectation of intended learning outcomes has a developmental aspect that includes the importance of considering impact on other people and society in general.

Some of our external constituents such as the business community propose affective components are as important in employment and citizenship settings as the cognitive components (Jones et al., 1995). This inclination to be open to evidence, to welcome new ideas, and to embrace complexity is as necessary or at least an equal partner to the pure cognitive component of reasoning skills (Facione et al., 1995). In some aspects of intellectual development, the cognitive and affective perspectives are inseparable. For example, tolerance has both cognitive and affective components: a rigidity of thinking and feeling (Erwin, 2000). It may be an artificial separation, but researchers (Facione et al., 1995) have separated the dispositional perspective, and more positive dispositions are expected due to the undergraduate experiences. Pascarella and Terenzini (2005) report the need for further research in intellectual development in studying college impact, particularly using longitudinal designs. To address this need, in part, this study focuses on the dispositional aspects of intellectual development.

Alumni Perceptions

Many institutions track some elements of intellectual and cognitive development during college, but what effects remain after college is also of fundamental interest.

Colleges and universities have surveyed their alumni since the 1930's (Pace, 1979), and the practice has grown in scope and consequence ever since. Information from alumni surveys is used to inform fundraising, college marketing, employment preparation, return on public investment, and institutional accountability such as accreditation, program review, and performance funding (Allen, Ramaekers, & van der Velden, 2005; Borden, 2005; Ewell, 2005; Volkwein, 2010). Over a dozen states conduct common comprehensive alumni surveys (Ewell, 2005), and several countries have incorporated the practice as well (Weerts & Vidal, 2005).

In spite of the ubiquitous nature of alumni surveys, questions are still posed about the credibility of self-report data (Bowman & Seifert, 2011; Pike, 2011). Nevertheless, as the foci of surveys have expanded from employment history to college outcomes, it is wise to keep in mind some possible limitations of survey methodologies as well as continue research into their validity. Some of the issues raised about survey methodology include the potential limits of generalizability due to low response rates, the possibility of other life events being confused with college impact, and the role of "indirect" measures versus "direct" measures of student development. In spite of these concerns, alumni perceptual data maintains a prominent place in institutional assessment portfolios.

Given the place of alumni information in institutions, it follows that information obtained during the undergraduate years be studied for its value in optimizing alumni perceptions. Before embarking on research studies with alumni, Volkwein (2010) recommends that researchers adopt an intended perspective. This study utilized the perspective of a developmental outcome model. That is, an intellectual developmental model based on dispositions was considered as the underlying expectation or educational objective for undergraduate impact. Stated another way, higher intellectual development should occur because of college. Generally, this study determines if average differences in college intellectual development occur between alumni satisfied or dissatisfied in nine areas of personal growth. These areas of personal growth are: the arts, creative thinking, making logical inferences based on assumptions, independent learning, setting personal goals, self-reliance, exercising initiative, persistence, and tolerating other points of view. It was hypothesized that satisfied alumni in their personal growth had higher intellectual development during college.

Method

Data Source and Participants

Samples of late sophomores (45-70 semester hours completed) enrolled during 1995-2000 at a moderately selective mid-sized institution in the mid-Atlantic area were selected as participants for this study. Students were chosen if the last digit of their identification number ended in seven. This sample was considered as quasi-random because no selection bias was associated with the last digit. Students in this sample were given a developmental test in a proctored setting. Although no comparison was made with other student groups (i.e. students whose identification did not end in seven) who completed other assessment instruments, no motivation bias was believed to be present with administration of this particular instrument to students with an id digit of seven.

Given the place of alumni information in institutions, it follows that information obtained during the undergraduate years be studied for its value in optimizing alumni perceptions.

Additional datasets were also obtained from institutional alumni surveys administered between 2002 and 2008. As part of the institution's ongoing assessment process, alumni surveys are administered every year to alumni who had graduated from a given group of major programs. Major programs participate in alumni surveys every six years, and graduates from all major programs at the institution are surveyed within five years. Institutionally, all graduates are contacted within one to five years after program completion. Approximately 40-50% of the graduates contacted responded to the survey, resulting in a sample matched with the Scale of Intellectual Development of complete data for this study of n=624 (See Table 1).

Table 1

Frequencies in Longitudinal Design of Late Sophomores Completing the Scale of Intellectual Development by Year of Alumni Survey Administration

Alumni Survey	Scale of Intellectual Development						
	1995	1996	1997	1998	1999	2000	Subtotals
2002	47	0	0	0	0	0	47
2003	63	81	4	0	0	0	148
2004	38	86	22	0	0	0	146
2005	29	28	22	18	36	0	133
2006	0	0	1	0	6	5	12
2007	0	0	21	12	26	22	81
2008	0	0	0	3	26	28	57
Totals	177	195	70	33	94	55	624

Instruments

The Scale of Intellectual Development (SID) (Erwin, 1983) was administered to a sample of sophomores who had completed 45-70 semester hours. Originally based on Perry's (1968) scheme of intellectual development , the SID has four subscales: Dualism, Relativism, Commitment, and Empathy. For the Dualism subscale, individuals scoring high on this factor tend to see issues in clear-cut, categorical terms, and look to authorities for the answers. For the Relativism subscale, individuals scoring high on this factor recognize alternative perspectives exist and can handle diversity within themselves and in relation to other people; but they still believe outside forces determine their future. For the Commitment subscale, individuals scoring high on this factor have begun to make major decisions in their lives and accept the responsibilities and consequences of these decisions. There is a tolerance of other viewpoints; however, these persons know where they stand in relation to other people. For the Empathy subscale, individuals scoring high on this factor not only have made major life decisions but also are aware of their impact on other people. High scorers have developed sensitivity about other people and feel responsible for improving society in general. Validity evidence for the SID as a measure of intellectual development may be found in Buczynski (1994), DeMars and Erwin (2003), Erwin (1993), and Erwin (2000).

Only Commitment and Empathy SID sub-scales were used in this study because these developmental dimensions are at the higher end of intellectual development and expected of college graduates. In addition, Commitment and Empathy were hypothesized to be related to the areas of personal growth collected from graduates. For the current sample, the subscales of Commitment and Empathy had Cronbach's alpha internal consistency coefficients of .80 and .86, respectively.

On the alumni survey, graduates were asked to indicate their employment and continuing education history; overall satisfaction with the institution, major program, and various student services; satisfaction with several academic areas such as writing, mathematics, speaking, and technology; and satisfaction with the institution's contribution to their personal growth in nine areas: the arts, creative thinking, making logical inferences based on assumptions, independent learning, selection of personal goals, self-reliance, exercising initiative, persistence, and tolerating other points of view. Alumni responded on a 4-point Likert scale ranging from *very dissatisfied* to *very satisfied* for each growth area. Responses were categorized in this study into either *satisfied* or *dissatisfied* for easier interpretability.

Procedure

The longitudinal sample in this study consisted of students who had completed the SID and later an alumni survey. The time between late sophomore developmental scores and post-graduate perceptions ranged from four to six years. The reason for this span was due to years spent until graduation and the target years the alumni survey was administered. Table 1 lists the frequencies related to year of cognitive development testing with year of completing alumni survey.

Some of our external constituents such as the business community propose affective components are as important in employment and citizenship settings as the cognitive components.

Multivariate analysis of variance (MANOVA) statistics were calculated using each of the nine alumni personal growth areas as independent variables and two subscales from the SID scale, Commitment and Empathy, as the dependent variables. MANOVA was used to study average differences in intellectual development assessed during college for satisfied or dissatisfied alumni in several areas of personal growth. The Commitment and Empathy subscales correlated 0.65; therefore, Roy's Maximum Root Criterion was chosen as the MANOVA statistic because it is most powerful when the dependent variables are concentrated in a single variate (Hair, Black, Babin, & Anderson, 2010). Cohen's *d* (Hartung, Knapp, & Sinha, 2008) was the effect size calculated in this study; *d* essentially reports the magnitude between groups in standard deviation units. Typically effect sizes of .2 or lower are considered small, .5 moderate, and .8 large differences. Cohen has noted that effect sizes in personal and social areas are likely to be smaller than in achievement tests.

Results

Table 2 reports the sophomore Commitment and Empathy means, Cohen's *d* statistics, and F statistics between the satisfied and dissatisfied alumni. MANOVA statistics are reported for each independent variable below.

Table 2

Means, Cohen's d, and Follow-Up ANOVA Statistics

SID Subscale	*F*	Satisfied (*n*)	Dissatisfied (*n*)	Cohen's *d*	*F*	Satisfied (*n*)	Dissatisfied (*n*)	Cohen's *d*
Appreciation of Arts	7.47**	62.6 (482)	54.80 (74)	0.33	17.57***	57.25 (482)	49.81 (74)	0.51
Creative Thinking	5.61**	62.43 (481)	55.58 (75)	0.29	7.13**	56.50 (481)	52.13 (75)	0.33
Logical Inferences	5.61**	62.39 (492)	55.00 (63)	0.31	9.78***	56.89 (492)	50.95 (63)	0.42
Learning Independently	8.17**	62.62 (486)	54.01 (68)	0.37	9.26**	56.92 (486)	51.33 (68)	0.39
Personal Goals	3.02	62.11 (506)	56.04 (49)	0.26	5.42*	56.72 (506)	51.77 (40)	0.35
Initiative	5.54**	62.28 (507)	54.00 (48)	0.30	11.06**	56.85 (507)	49.75 (48)	0.50
Other Points of View	7.28**	62.58 (446)	55.81 (105)	0.29	9.09***	56.91 (446)	52.30 (105)	0.33

*$p<.05$, **$p<.01$, ***$p<.001$

In general, graduates "satisfied" with their education in the arts, creative thinking, logical inferences, learning independently, exercising initiative, and tolerating other points of view had higher Commitment and Empathy scores as sophomores than "dissatisfied" graduates in these areas of personal growth. Graduates satisfied with their ability to set personal goals had higher average Empathy scores than dissatisfied graduates. No differences were found in either sophomore Commitment or Empathy scores for the personal growth areas of self-reliance or persistence.

Also reported in Table 2 are Cohen's *d* statistics across the various personal growth areas. These were small, ranging from a low of 0.29 for Commitment in creative thinking to a high of 0.51 for Empathy in the arts.

Appreciation of Arts

The MANOVA overall statistic of Roy's Maximum Root Criterion was 0.03, $F(2, 555)=9.22$, $p=.0001$ for the alumni survey question about developing appreciation in the arts. Analysis of variance follow-up results for Commitment was $F(1, 556 =7.47$, $p=.006$; and for Empathy was $F(1, 556)=17.57$, $p=.0001$. "Satisfied" Arts Commitment mean of 62.78 (n = 484) was greater than the "dissatisfied" Commitment mean of 54.81 (n=74). The "satisfied" Arts Empathy mean of 57.25 (n = 484) was greater than the "dissatisfied" Empathy mean of 49.81 (n = 74).

Creative Thinking

Roy's Maximum Root Criterion was 0.01, $F(2, 555) = 3.79$, $p = .02$ for the alumni question pertaining to creative thinking. Analysis of variance follow-up results for Commitment was $F(1, 556) = 5.85$, $p = .01$; and Empathy was $F(1, 556) = 7.42$, $p = .006$. Graduates "satisfied" with creative thinking had a Commitment mean 62.61 (n = 483), which was greater than "dissatisfied" Commitment mean of 55.58. Graduates "satisfied" with their creative thinking had a higher Empathy average of 56.96 (n = 483) than the "dissatisfied" average of 52.13 (n = 75).

Logical Inference

Roy's Maximum Root Criterion was 0.01, $F(2, 554) = 5.03$, $p = .006$ for the alumni survey question about making logical inferences. ANOVA follow-up for Commitment was $F(1, 555) = 5.83$, $p = .01$; and Empathy was $F(1, 555) = 10.06$, $p = .001$. Graduates "satisfied" with their education in developing logical inference had Commitment means of 62.56 (n = 494), which was greater than "dissatisfied" alumni having a mean of 55.00 (n = 63).

Learning Independently

Roy's Maximum Root Criterion was 0.01, $F(2, 553) = 5.04$, $p = .006$ for the alumni survey question about Learning Independently. ANOVA follow-up for Commitment was $F(1, 554) = 8.43$, $p = .003$; and Empathy was $F(1, 554) = 9.56$, $p = .002$. Commitment mean for graduates "satisfied" with their education about Learning Independently was 62.79 (n = 488), which was greater than the mean of "dissatisfied" alumni of 54.01 (n = 68). Empathy mean for "satisfied" alumni was 57.0 (n = 488), which was higher than the "dissatisfied" mean of 51.33 (n = 68).

Personal Goals

Roy's Maximum Root Criterion was 0.01, $F(2, 554) = 4.96$, $p = .05$. ANOVA follow-up for Commitment was $F(1, 555) = 3.16$, $p = .07$; and Empathy was $F(1, 555) = 5.61$, $p = 0.01$. No statistical difference was found on Commitment between graduates "satisfied" with their education about their Personal Goal development and "dissatisfied" alumni. The Empathy mean for "satisfied" alumni was 56.80 (n = 508), which was higher than the "dissatisfied" mean of 51.78 (n = 49).

Self-Reliance

Roy's Maximum Root Criterion was 0.009, $F(2, 555) = 2.55$, p=.07 for the alumni question pertaining to Developing Self-Reliance. No follow-up ANOVA were reviewed because this overall MANOVA p value of .07 was greater than the preset alpha level of .01. Therefore, no differences were found on Commitment or Empathy between alumni "satisfied" versus "dissatisfied" with their prior collegiate experience in Developing Self-reliance.

Initiative

Roy's Maximum Root Criterion was 0.02, $F(2, 554) = 5.74$, $p = .003$. ANOVA follow-up for Commitment was $F(1, 555) = 5.72$, $p = .01$; and Empathy was $F(1, 555) = 11.29$, $p = .0008$. Commitment mean for graduates "satisfied" with their education developing their capacity for Exercising Initiative was 62.45 (n = 509), which was greater than the mean of "dissatisfied" alumni of 54.00 (n = 48). Empathy mean for "satisfied" alumni was 56.96 (n = 509), which was higher than the "dissatisfied" mean of 49.75 (n = 48).

Persistence

Roy's Maximum Root Criterion was 0.004, $F(2, 555) = 1.25$, $p = .28$. This p value of .28 was greater than the preset alpha of .01, and no follow-up ANOVA were reviewed. No differences were found between alumni "satisfied" versus "dissatisfied" in their prior collegiate experience of Developing Persistence on either Commitment or Empathy.

Tolerating Other Points of View

Roy's Maximum Root Criterion was 0.01, $F(2, 550) = 4.89$, $p = .007$. ANOVA follow-up for Commitment was $F(1, 551) = 7.62$, $p = .006$; and Empathy was $F(1, 551) = 9.50$, $p = .002$. Commitment mean for graduates "satisfied" with their education for Tolerating Other Points of View was 62.76 (n = 448), which was greater than the mean of "dissatisfied" alumni of 55.82 (n = 105). Empathy mean for "satisfied" alumni was 57.05 (n = 448), which was higher than the "dissatisfied" mean of 52.03 (n = 448).

Discussion

Despite…limitations, this study offers the strength of a longitudinal study and the collection of a "direct" measure of intellectual development.

Graduates who responded to an alumni survey about their satisfaction with several areas of personal growth were linked with their sophomore intellectual development scores, as measured by the SID, four to six years prior. Unlike most previous studies that focus either on current students or alumni, this study examines student development over time. "Satisfied" graduates in their abilities with the arts, creative thinking, logical inferences, learning independently, exercising of initiative, and tolerating other points of view had higher Commitment and Empathy scores on the average than "dissatisfied" graduates in these areas of personal growth. Graduates satisfied with their education in personal goals had higher Empathy average scores than dissatisfied graduates. No differences in intellectual development were found between "satisfied" and "dissatisfied" graduates in self-reliance or developing persistence. Personal goals had mixed results; no differences were found for Commitment, but satisfied graduates had higher Empathy scores. While it is not known why differences in intellectual development were not found in these areas, greater intellectual development in college may not be expected to be related to all these areas of personal growth. Self-reliance and persistence may be related to other areas of college student development but not intellectual development.

These results are limited to graduates who responded to the institutional alumni survey and who had completed the SID, which was administered to quasi-random samples of late sophomores four to six years earlier. Approximately half of the graduates who had taken the intellectual development test responded to the alumni survey. It is not known if the non-responding group of alumni would change these results had they responded. "Satisfaction" was defined and reported by the graduates themselves that might vary by self-referencing differences on the personal growth dimensions, and by a greater proportion

of graduates who might acquiesce or be "satisfied" with institution. Despite the average differences found and reported in several personal growth areas, the effect sizes using Cohen's *d* were moderate. Some of the differences may be due to larger sample sizes, and future studies should continue to monitor effect size. On the other hand, the effect sizes in this study are similar to freshmen to senior differences reported in the low .30's for the major proprietary general education tests of Collegiate Learning Assessment, the Collegiate Assessment of Academic Proficiency, and the Proficiency Profile (Klein et al., 2009).

Despite these possible limitations, this study offers the strength of a longitudinal study and the collection of a "direct" measure of intellectual development. Although this study did not seek to identify what variables were associated with higher or lower intellectual development scores during college, prior research (e.g. Pascarella & Terenzini, 2005) has shown that offering academic and student services varying in perspective and opinions helps students reach commitments. In addition, students' empathy can be enhanced by having them consider the effects of their and others' decisions on society and humankind in general. Both Commitment and Empathy are currently supported in concept in the Degree Qualifications Profile. Whatever the reasons for students being higher in Commitment and Empathy as undergraduates, graduates satisfied with several personal growth dimensions are associated with higher commitment and empathy scores from their earlier college years.

Studying possible connections of collegiate student development with personal growth has implications for programming at the undergraduate level as well as our understanding of the lasting impact of development after college

This study also supports the importance of intellectual development on graduates' perception about their own collegiate experience. The concept of general education and generic intellectual development skills is challenged by voices wishing to shorten the length of the undergraduate period either by eliminating general education entirely or to focus just on the major or professional education. This study also supports the value of graduates' self-ratings of several areas of personal growth; satisfied graduates also had higher intellectual development scores in their college years. Often alumni perceptions are questioned as to their value: are they isolated or are they important as revelations to earlier collegiate development? Causal links are not made here, but relationships between alumni growth with earlier intellectual development was encouraging. This relationship supports college impact and value of alumni perceptions.

Future studies may utilize other conceptualizations of intellectual development and other alumni personal growth areas. Retesting in intellectual development during the undergraduate period and also in direct testing of alumni in intellectual development may also enhance, or cast other perspectives on these findings. Other areas of student development besides intellectual development may be used such as moral or psychosocial development. Using more content-related measures of general education such as scientific reasoning or communication abilities might be examined too. The undergraduate experience is complex, and additional dimensions of general education would be useful to study.

Nevertheless, this study encourages future work in intellectual development during the collegiate experience and in continued reliance on alumni perceptions. Longitudinal studies of this type are also desirable given the current emphasis on lifelong learning. The link of association in this study supports further activities in both areas of intellectual development and in alumni information.

References

Adelman, C., Ewell, P., Gaston, P., & Schneider, C. G. (2011). *The Degree Qualifications Profile.* Indianapolis, IN: Lumina Foundation for Education. Retrieved from: http://www.luminafoundation.org/publications/The_Degree_Qualifications_Profile.pdf

Allen, J., Ramaekers, G., & van der Velden, R. (2005). Measuring competencies of higher education graduates. *New Directions for Institutional Research, 126*, 49-59.

Bologna Process, Retrieved April 9, 2012 from http://www.ond.vlaanderen.be/hogeronderwijs/bologna/about/.

Borden, V. M. (2005). Using alumni research to align program improvement with institutional accountability. *New Directions for Institutional Research*, *126*, 61-72.

Bowman, N. A., & Seifert, T. A. (2011). Can college students accurately assess what affects their learning and development? *Journal of College Student Development, 52*(3), 270-290.

Buczynski, P. L. (1994). Using hierarchical modeling and practical measures of fit to evaluate structural equation models. *Measurement and Evaluation in Counseling and Development, 26*, 328-338.

Collins, R. A. (2006). Cognitive development of adult undergraduate students: Cohort and non-cohort settings. *Educational Considerations, 33*, 35-41.

Dawson, T. (2004). Assessing intellectual development: Three approaches, one sequence. *Journal of Adult Development, 11,* 71–85.

DeMars, C., & Erwin, T. D. (2003). The Scale of Intellectual Development: Application of an unfolding model. *Journal of College Student Development*, *44*(2), 168-184.

Erwin, T. D. (1983). The Scale of Intellectual Development: Measuring Perry's scheme. *Journal of College Student Personnel, 24*, 6-12.

Erwin, T. D. (2000). *The NPEC sourcebook on assessment, volume 1: Definitions and assessment methods for critical thinking, problem solving, and writing.* U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Ewell, P. T. (2005). Alumni studies as instruments of public policy: The U.S. experience. *New Directions for Institutional Research, 126*, 19-29.

Evans, N. J., Forney, D. S., Guido, F. M., Patton, L. D., & Renn, K. A. (2010). *Student development in college: Theory, research, and practice* (2nd ed.). San Francisco, CA: Jossey-Bass.

Facione, P., Sanchez, C., Facione, N., & Gainen, J. (1995). The disposition toward critical thinking. *Journal of General Education*, *44*, 1-25.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). New York, NY: Prentice-Hall.

Hartung, J., Knapp, G., & Sinha, B. K. (2008). *Statistical meta-analysis with application.* Hoboken, NJ: Wiley.

Jones, E., Hoffman, S., Moore, L., Ratcliff, G., Tibbetts, S., & Click, B. (1995). *National assessment of college student learning: Identifying college graduates essential skills in writing, speech, and critical thinking* (Research and Development Report No. NCES 95-001). Washington, DC: U. S. Government Printing Office.

Kenny, D. A. (1987). *Statistics for the social and behavioral sciences.* Boston, MA: Little, Brown.

Klein, S., Liu, O.L., Sconing, J., Bolus, R., Bridgeman, B., Kugelmass, H., Nemeth, A., Robbins, S., & Steedle, J. (2009, September 29). *Test validity study (TVS) Report.* Supported by the Fund for Improvement of Postsecondary Education (FIPSE). Online at: http://www.voluntarysystem.org/index.cfm?page=research.

Pace, C. R. (1979). *Measuring outcomes of college: Fifty years of findings and recommendations for the future.* San Francisco, CA: Jossey-Bass.

Pascarella, E. T. (1985). College environmental influences on learning and cognitive development: A critical review and analysis. In J. C. Smart (Ed.), *Higher education: Handbook of theory and research* (pp. 1-64). New York, NY: Agathon Press.

Pascarella, E. T., & Terenzini, P. T. (2005). *How college affects students: A third decade of research* (2nd ed.). San Francisco, CA: Jossey-Bass.

Perry, W. G., Jr. (1968). *Forms of intellectual and ethical development in the college years: A scheme.* New York, NY: Holt, Reinhart, and Winston.

Pike, G. (2011). Using college students' self-reported learning outcomes in scholarly research. *New Directions for Instittional Research, 150*, 41-58.

Taube, K. (1997). Critical thinking ability and disposition as factors of performance on a written critical thinking test. *Journal of General Education, 46*, 129-164.

Torres, V. (2003). Influences on their ethnic identity development of Latino college students in the first two years of college. *Journal of College Student Development, 44*, 532-547.

Volkwein, J. F. (2010). Assessing alumni outcomes. *New Directions for Institutional Research, S1*, 125-129.

Wang, Y., & Rodgers, R. (2006). Impact of service-learning and social justice education college students' cognitive development. *NASPA Journal, 43*, 316-337.

Weerts, C., & Vidal, J. (2005). Enhancing alumni research: European and American perspectives. *New Directions for Institutional Research, 126*, 49-59.

Book Review

Good Education in an Age of Measurement.
Gert J.J. Biesta. Boulder, CO: Paradigm Publishers, 2010.
160 pp. ISBN-13: 978-1594517914. Paperback, $29.95.

REVIEWED BY:
Brian W. Lagotte, Ph.D.
University of Kansas

In *Good Education in an Age of Measurement,* Gert J.J. Biesta argues that analysis about what constitutes a "good" education demands more than the evidence-based, "best practice" paradigm currently offers. Furthermore, the narrow perspective of assessing learning outcomes may prove detrimental for education towards a deeply democratic society. Although not exactly the type of insight assessment researchers might welcome, Biesta's thoughtful critique can ultimately enhance the ways scholars evaluate the quality of education. Biesta reinvigorates discussions about what constitutes a good education, specifically the purpose of education. Concerned about a lack of attention to purposes in the research literature, Biesta puts this issue front and center. His inquiry includes a normative perspective rather than only a managerial focus on education as a technique. That is, he produces a conceptual framework for why we ought to focus on particular educational goals. To this end, Biesta provides a three-prong framework for education, which should highlight a distinct outcome: producing a deliberative democratic order of increasing equity.

Although not exactly the type of insight assessment researchers might welcome, Biesta's thoughtful critique can ultimately enhance the ways scholars evaluate the quality of education.

Although perhaps not intentionally, the argument is usefully split in two parts: the first half of the book shows why employing only the evidence-based paradigm is inadequate for evaluating good education; the second half delves more directly into the philosophy of education to propose a pedagogy of interruption. Assessment researchers may find the first half a bit harsh, even though the material has been fairly well explored by sociologists, anthropologists, and educators grounded in the critical tradition who focus on the politics of curriculum and schooling (e.g. Apple, 1982, 2000). The critique is not, however, an attack on evidence-based practitioners but an analysis of the consequences of overemphasizing the assessment of learning outcomes ("learnification" in Biesta's words) to determine quality education. In the second half, Biesta turns from the political-economic analysis of education research to philosophy proper, proposing key elements of a good education. The focus turns to educating youth to promote a deep deliberative democracy—not just joining a democratic order already established, but constantly challenging the arrangement to be more equitable. As I summarize these points below, I pay more attention to the front end since the readers of *Research & Practice in Assessment* may be directly implicated.

The critique is not, however, an attack on evidence-based practitioners but an analysis of the consequences of overemphasizing the assessment of learning outcomes ("learnification" in Biesta's words) to determine quality education.

In Chapter One, Biesta argues that quantitative, data-driven measurements require advanced technique, and a fixation on the technical questions of assessment obscures more important normative questions about what exactly we ought to be measuring. As Biesta argues, "This has to do with the question of whether we are indeed measuring what we value, or whether we are just measuring what we can easily measure and thus end up valuing what we (can) measure" (Biesta, 2010, p. 13). In other words, it is fairly easy to measure the change in test scores of a school from one year to the next and determine if that school is making adequate progress. Suddenly, the ability to raise students' test scores in a standardized format, focusing mainly on written and mathematical literacy, determines a quality education. So, in my example, while schools are scrambling to capitalize on tutoring services, cutting "frivolous" classes like the arts and music, and drilling test-taking strategies in classes to capture the coveted title of a "high performing school," Biesta is merely asking, "Exactly why ought we train youth to ace standardized tests?"

My suspicion, like Biesta's, is that this question is not asked because evidence-based experts feel it has been answered. "Common sense" is that schooling qualifies children for a job or college. Therefore, educators gather evidence to assess how well students are learning the skills necessary for these paths. However, Biesta reminds, "We shouldn't forget, of course, that what appears or presents itself as 'common sense' often serves the interests of a particular group much better than the interests of other groups" (p. 15; see also Gramsci, 1971). When educators reduce schooling to qualification for the job market, or university in preparation for a higher skilled job market, education serves the interests of the current economic sector of society, in turn neglecting other important educational values such as citizen education or the arts (e.g. Nussbaum, 2011). But, Biesta argues qualification is merely one of three dimensions of education.

Biesta proposes that the three functions education can/ought to perform are qualification, socialization, and subjectification (pp. 19-21). Qualification, for example

job training, is providing youth knowledge and skills to do something. Socialization, whether explicit or implicit, integrates individuals into particular social structures. The last function, subjectification, is a far more slippery concept, but the crux of the purpose of a good education that comes later in the book. In short, subjectification promotes autonomous actors who are critically, creative, independent thinkers. Furthermore, subjectification creates individuals who do not merely fit into the social order as is, but who are capable of altering the status quo to bring about social arrangements of greater equity and deeper democracy. Finally, education is a composite of the three functions, impossible to separate (if effective). According to Biesta, subjectification is receiving less attention in discussions about effective education.

In Chapters Two and Three, Biesta makes his strongest points against the technocratic model of learning assessment. This model concentrates too heavily on the efficiency of education techniques for transmitting knowledge, but without questioning the knowledge that is being efficiently transmitted. In other words, "the question that always needs to be asked is 'Effective for what?'" (p. 34). Biesta also asks the politically important question of "effective for whom?" Too much time is spent in laboratory settings, randomizing controlled trials to quantify correlating phenomena; too little time is spent connecting these results with the contemporary social context. According to Biesta, "A key problem with the idea of evidence-based practice is that it simply overlooks the cultural option" (p. 45). So, the evidence-based assessment frame focuses on the technical aspect of education without paying equal attention to the moral elements. On the surface, technocratic solutions can mean that teachers become effective at increasing student learning, but rarely are they equipped to question why students are learning. On a deeper level, emphasis on technique means that education takes on a much more reproductive function and loses the potential to be a transformative social institution.

Biesta is merely asking, "Exactly why ought we train youth to ace standardized tests?"

Research concentrating only on "what works," Biesta argues, is insufficient to evaluate the quality of education in society—both practically and politically. The phrase "what works" seems to mean that educational techniques have been rigorously tested to discover the ideal practices; teachers can go to clearinghouses to collect these tools for their classroom, and the tools "will work" in each particular context. But, that is not exactly what is happening. Technically, a particular teaching practice shows a statistically significant relation to a specific consequence (controlling for a range of effects), said practice "has worked" under those conditions, thus it is deemed to be the thing that works under all "similar" conditions in the real world (p. 44). Nothing about the latter process contains the important logical connector "and it will work for your specific context every time." Politically, without a normative foundation of why teachers ought to use these techniques, emphasizing "what works" stifles critical decision-making by individual teachers. It "denies educational practitioners the right not to act according to evidence about 'what works' if they judge such a line of action would be educationally undesirable" (emphasis in original, p. 47). If teachers deem evidence-based practices ill fit to their own circumstances, they risk criticism about ignoring "what works," potentially deprofessionalizing their roles as creative educators. Thus, both processes, recapitulating what works and closing off the options for individual creativity, increase the tendency of educational institutions to reproduce the status quo rather than interrupt common sense understandings (e.g. Bourdieu & Passeron, 1977).

A good education prepares students with the tools to generate a deliberative democracy with the critical capacity to interrupt the status quo for increased social equity.

Unfortunately, one current understanding of schooling circulating is the educational field as marketplace, and in Chapter Three Biesta shows that assessment and accountability regimes perpetuate this frame. He argues that the "accountability" narrative is becoming an overwhelming logic in which customers (parents and students) can hold service providers (teachers and schools) accountable for an excellent product (education). It should not be surprising, therefore, that parents with money can afford a better education, reproducing economic inequality between generations. When Biesta reminds us that contemporary discourses express neoliberal themes, especially when discussing parental choice, he hints at another influence of assessment and accountability (pp. 55-59). I would argue that if parents (and students) are consumers in a competitive market, they need a clear measurement of quality among (increasingly charter) schools to make their purchasing decisions. So, the State intervenes to mandate a curriculum that is standard across schools, thus comparable (e.g. Common Core). Assessment researchers determine which schools are most effective in getting students to learn that curriculum—making yearly progress adequately—and media publish "league tables." In this sense, assessment research is the academic version of Consumer Reports for schools in an era of neoliberalism.

After the first half of the book, where Biesta is making a pretty straightforward critique on the consequences of contemporary education research, he sets out his argument of items to consider for a "good education" from the end of Chapter Three to Chapter Six. For readers unfamiliar with the philosophy of education in general, or the work of Zygmunt Bauman, Hanna Arendt, Jurgen

Oelkers, and Jacques Ranciere specifically, the last half of the book might take a bit more effort. For now, I will paraphrase (what I understand as) the main theme of Biesta's argument and leave the longer philosophical explorations to the reader. The work required to unpack the complexity is both rewarding and substantially aided by the crisp conclusions at the end of each chapter; it may be wise to read the conclusions of the final three chapters before the content itself.

Biesta is asking education researchers, qualitative and quantitative alike, to provide theoretical, empirical, and normative justification for the choice of any quality measure.

A good education prepares students with the tools to generate a deliberative democracy with the critical capacity to interrupt the status quo order for increased social equity. The focus, to return to the first chapter, is to reinvigorate the subjectification function of education so students are not merely instructed on the techniques of a stagnant democratic order. "The pedagogy of interruption thus has its place in the domain of subjectification" (p. 91) because qualification and socialization are activities that function to prepare individuals to fit into a system as is. Education, through the third element, is responsible "for coming into presence of unique individual beings…for the plurality that is the condition of human action and human freedom" (p. 91). Therefore, a key measure of good education for Biesta is how it increases freedom for all individuals, which requires both the knowledge about how to deliberate towards a deep democracy and the responsibility to make it happen. While qualification and socialization can provide much of the former, subjectification must provide much of the latter. And, assessing learning outcomes is not designed to address the question of subjectification in education. Thus, the evidence-based "what works" paradigm cannot capture all aspects of what makes a good education.

So, why would assessment scholars, readers of RPA, bother to pick up a book by a philosopher of education who specifically criticizes the exact paradigm within which they work? Researchers should read it because it is both provocative and challenging. One cannot engage with the literature that Biesta draws from and simply dismiss his argument as misguided or utopian because he is a philosopher rather than an assessment specialist. Biesta is asking education researchers, qualitative and quantitative alike, to provide the theoretical, empirical, and normative justification for the choice of any quality measure. Furthermore, he is demanding an equally grounded answer to what we ought to be teaching and measuring. The current evidence-based paradigm is not fully equipped to answer these questions; "what works" is best at answering "how" to teach and measure, not "why." While important, technique is a means not an end; thus, Biesta remains skeptical that assessment alone can evaluate a "good" education.

References

Apple, M. W. (1982). *Education and power*. London: Routledge.

Apple, M. W. (2000). *Official knowledge: Democratic education in a conservative age*. New York, NY: Routledge.

Biesta, G. J. J. (2010). *Good education in an age of measurement: Ethics, politics, and democracy*. Boulder, CO: Paradigm Publishers.

Bourdieu, P., & Passeron, J. C. (1977). *Reproduction in education, society, and culture*. London: Sage Publications.

Gramsci, A. (1971). *Selections from the prison notebooks of Antonio Gramsci*. New York, NY: International Publishers.

Nussbaum, M. C. (2011). *Creating capabilities: The human development approach*. Cambridge, MA: Belknap Press of Harvard University Press.

Book Review

Science Learning and Instruction: Taking Advantage of Technology to Promote Knowledge Integration. Marcia C. Linn and Bat Sheva Eylon. New York, NY: Routledge, 2011. 360 pp. ISBN-13: 978-080586055X. Paperback, $39.95.

REVIEWED BY:
Jabari Mahiri, Ph.D.
University of California, Berkeley

In *Science Learning and Instruction: Taking Advantage of Technology to Promote Knowledge Integration,* Linn and Eylon make a critical shift in the focus of assessment. In the quest to assess what students learn, they show why we must also assess how they learn. These researchers argue that this approach to assessment can substantively increase the quality of student knowledge when embedded in a process called "knowledge integration" (KI). They also demonstrate how KI can improve the quality of science learning overall when instruction, assessment, professional development, and school leadership are systematically aligned. These ameliorative possibilities begin with a simple premise: "Everyone can learn science" (p. ix).

In daily life everyone observes, predicts, reflects upon, and has intuitive beliefs and ideas about things that occur in the natural world. Linn and Eylon illustrate how science instruction can effectively build on these experience-based understandings of the natural world. They show how learning can be guided through processes of interrogating information, gathering evidence, and using technology to help students elicit, add to, distinguish, sort, reflect on, and evaluate scientific ideas. They see assessment within this process as not just descriptive of what students learn, but as generative of how they learn. When students are taught and assessed in this manner, a foundation is established for them to successfully employ scientific perspectives to understand the natural world and to become life-long science learners.

Science Learning and Instruction: Taking Advantage of Technology to Promote Knowledge Integration counters the notion in education that transmitting information is central to learning. The authors call this widely accepted notion "the absorption approach." Its hallmarks in science instruction require students to listen to lectures, read textbooks and complete exercises, and conduct prescribed experiments or investigations following step-by-step procedures. Fundamental aspects of the absorption approach do not directly build on the interests and intuitions of learners.

In contrast, Linn and Eylon's knowledge integration approach offers a different process that focuses on learning as increasing abilities to make sense of increasingly complex ideas. In their approach information transmission is supplanted by idea generation, and the process begins with the learner's own ideas. So, the concept of "ideas" is core to KI. They define an idea as "each distinct view held by the learner. Ideas include observational, intuitive, mathematical, visual, and analogical views" (p. 26).

Students hold many ideas and beliefs about a range of topics including science topics, but these intuitive views are usually not scientifically accurate. But they are important starting points for developing understanding of science. The authors note, "By eliciting ideas instruction takes advantage of the diverse, culturally-determined views students develop and ensures that all students can build on their views" (p. 26). The ability for instruction to connect science topics to previously held views of learners is crucial for students to take on identities as science learners. One value of the KI approach is that it methodically develops and complicates student ideas in ways that are reflective of how scientists research and solve problems and increase their understanding of the natural world.

Linn and Eylon make a critical shift in the focus of assessment. In the quest to assess what students learn, they show why we must also assess how they learn.

The KI approach stimulates and guides the development of a repertoire of ideas by surfacing the views already held by the learner. It facilitates learners adding new ideas to those previously held based on their backgrounds, experiences, personal interests, and cultural frames. Through instructional activities, students work to generate, distinguish, sort, and reflect on ideas and use multiple levels of analysis to discover emerging and resilient patterns or connections. Eventually, students are enabled to evaluate the significance and accuracy of emerging ideas using increasingly nuanced, evidence-based criteria. This process for achieving coherent, scientific understanding is what the authors mean by knowledge integration.

Throughout the book instruction and assessment practices that use the knowledge integration approach for teaching a range of science topics are described, analyzed, and contrasted to the absorption approach in order to illustrate the advantages of KI. Case studies of curricular and pedagogical designs for teaching thermodynamics (Chapters 3) and the particulate structure of matter (Chapter 4) reveal definitive KI advantages for learning in these domains while also illustrating how the KI foundation and framework emerged.

Each chapter addresses ways to align assessment with instruction, the advantages of technology for enabling KI, and specific roles for teachers in guiding students' knowledge integration. "Reflection Activities" are also provided at the end of each chapter to encourage further consideration, discussion, and understanding of the topics

and issues covered as well as to promote lifelong science learning. A major benefit for educators, researchers, and software designers is that most of the science curricula discussed are available for free as Web-based Inquiry Science Environment units (complete with open source authoring tools) at WISE.Berkeley.edu.

The authors' considerations of the value of visualizations threads throughout the book and offers a focal point for discussing how their compelling approach to assessment connects to evolving roles for teachers and technology in guiding student learning. They describe the significance of visualizations with regard to the success of instruction in both case study chapters (3 and 4), and they further explicate its value in Chapter 8. Teachers can utilize technology to support high levels of science learning because it helps generate and add to the complexity of ideas about things in the natural world that cannot be directly observed.

This approach to assessment can substantively increase the quality of student knowledge when embedded in a process called "knowledge integration."

Essentially, modern, on-line learning environments enable more sophisticated scientific visualizations. For example, they allow learners to explore chemical or molecular reactions that are too small, or too massive, or that occur too rapidly to be apprehended by the naked eye. They also allow interactive manipulations in connection with these phenomena like user customizations, prototyping, simulations, virtual experiments, and other self-directed explorations. Importantly, the authors provide extensive examples of how technology aids teachers in embedding assessments directly and seamlessly in the actual process of student learning.

Learning assessments enhanced by visualizations can be embedded within each stage of the KI process. But as the authors indicate, "Matching visualizations to the knowledge levels of students is essential for the visualizations to succeed" (p. 207). In chapter 8 they present findings and insights from numerous studies that show the efficacy of visualizations for helping students elicit, add to, distinguish among, sort out, and reflect upon scientific ideas. In making complex thinking visible at each KI stage, opportunities for assessing intricate and discreet aspects of learning are enhanced in ways that clearly extend the quality of student learning overall.

Linn and Elyon also describe how student writing (and drawing) can be incorporated into the stages of KI as additional ways to assess and make student learning visible. For example, they report that in a thermodynamics unit students were asked to write critiques of responses to how well a thermal equilibrium experiment was conducted. The critiques were coded and scored using a KI rubric and compared to responses of another group of students who watched a visualization of a similar experiment as well as to responses of a third student group that conducted their own virtual experiments. The assessment and scoring rubric also included an item for all three groups of students to draw and explain pictures to show how heat is transferred. Based on pretests and posttests, outcomes were that all three groups made significant gains on KI questions although the critique group outperformed the other two. Ultimately, the results suggest that a range of visualization activities and strategies help students develop more coherent understanding of complex scientific phenomena.

The authors also critique uses of visualizations in the absorption approach and argue that these uses are often problematic. For example, the absorption approach sometimes uses visualization in attempts to motivate or engage students in learning. But without nuanced designs and guidance of learning as shown in the KI approach, visualizations aimed at transmitting information have been found to be inappropriate, misaligned, or overly complicated in terms of interpretation demands for achieving systematic science learning. In this regard the authors note, "Visualizations add value when they align with the knowledge of the learner, focus on difficult ideas that are hard to express in other formats, take advantage of interactivity, and support self-directed explorations for science topics that are fundamentally dynamic" (p.27).

Is the KI approach applicable for improving student learning beyond science topics? Yes. In fall semester of 2011, I used *Science Learning and Instruction: Taking Advantage of Technology to Promote Knowledge Integration* in a graduate class on urban education for a diverse group of pre-service elementary teachers and secondary science, math, and English teachers. They read, discussed, and presented on chapters in the book and also explored some of the web-based resources. But I initially introduced the knowledge integration process by modeling a literature lesson using a canonical poem by Robert Hayden entitled "The Whipping."

In their approach information transmission is supplanted by idea generation, and the process begins with the learner's own ideas.

After a volunteer read the poem to the class, students were asked to write their ideas about what happened on their paper copies. Afterwards, volunteers were asked to share their ideas with the class as I recorded them on the board without making evaluations. Students were asked to add more ideas to those on the board. Then working in groups, they discussed ways to distinguish and sort the ideas. After further reflection they evaluated

which ideas provided the most coherent and convincing meaning(s) of the poem based on their evolving understanding. As the lesson concluded, students saw the advantages of generating, adding, distinguishing, evaluating, and eventually integrating a wealth of ideas about a complex written text. Rather than transmitting a preconceived meaning of the poem, students were guided to develop and collaborate on a range of coherent and defensible meanings with clear opportunities for embedded assessments of learning at each stage.

Rather than transmitting a preconceived meaning of the poem, students were guided to develop and collaborate on a range of coherent and defensible meanings with clear opportunities for embedded assessments of learning at each stage.

Linn and Eylon's use of writing for assessing learning within the KI process also connects to its value for assessment beyond science learning. Elements of their methods for using problem-based writing prompts with automated scoring rubrics can be appropriated across a range of disciplines to increase the effectiveness of teachers in continually assessing and guiding students in the actual process of learning. Overall, this book provides clear direction for the transformation of science learning and instruction, and it has important implications for teaching and learning in other disciplines as well.

Software Review

jMetrik item analysis [software application]. Patrick Meyer. Retrieved from http://www.itemanalysis.com/

REVIEWED BY:
Andrea Gotzmann, Ph.D.
Medical Council of Canada

Louise M. Bahry, M.Ed.
University of Massachusetts, Amherst

Technology, and the use of software to enhance or assist with evaluating measurement statistics, is currently a large emphasis for users. Measurement statistics, used in classical test theory (CTT) and item response theory (IRT), have been elusive for some users, as the measurement concepts are complex and investment of time to understand is intensive (Lord, 1980; Lord & Novick, 1968). However, users across many content disciplines are developing their understanding and applying these methodologies to new areas (i.e., medical education, psychology, etc.). As a result, the needs of researchers and applied practitioners have changed, and consequently, require tools to apply psychometrics. Reliance on specialized or esoteric software has been the norm; however, according to Drasgow, Luecht, and Bennett (2006), "Technology offers solutions to many of the challenges faced by testing programs" (p. 471). That is, technology may provide many of the psychometric analyses to be more accessible to broader audiences, so that users of all levels of expertise can take advantage of the advances in educational measurement.

Technology may provide many of the psychometric analyses to be more accessible to broader audiences, so that users of all levels of expertise can take advantage of the advances in educational measurement.

There are many statistics that are calculated and evaluated for both CTT and IRT, and many of them housed in separate programs. Some testing organizations create proprietary software to integrate these analyses, or create user defined solutions with programs such as R, SAS or SPSS. This review describes and evaluates a new software program, called jMetrik, version 2.1.0, that can produce psychometric statistics for both CTT and IRT. jMetrik allows for a more integrated system to conduct psychometric analyses for research and operational purposes without the cost associated with some other psychometric software programs.

Program Description

The jMetrik software was developed by J. Patrick Meyer, Ph.D., with support from the National Council on Measurement in Education (NCME), as a recipient of the Bradley Hanson Award for Contributions to Educational Measurement in 2010. jMetrik is a free and open source software application for classical and modern psychometric analyses. The program is a pure Java application that runs on Windows, Mac, OSX, and Linux platforms, with requirements of 256 MB of available memory, and Java 6 (i.e., JRE 1.6) or higher. The jMetrik graphical user interface (GUI) combines a workspace tree, data view, point-and-click menu, and several dialog boxes. Although the software is currently available not all features are active, or fully functional. Therefore, this review will address the features that currently are available, offering a snapshot of the current version of the software.

Current Available Analyses and Program Interface

The jMetrik software includes psychometric analyses such as CTT, IRT, Differential item functioning (DIF), and Confirmatory Factor Analysis (CFA). All of these analyses are useful in evaluating the psychometric quality of an assessment. In addition, the program offers many graphing features such as histograms and nonparametric IRT characteristic curve estimation.

The data interface to import data is similar to Microsoft Access. Data cannot be entered directly into the program; to import, the data need to be structured as a comma or space delimited text file, where missing values must represented by two consecutive commas or spaces and missing data is automatically scored as zero. The user needs to score the data and indicate the score key for each item as well as missing data; this step is required for both scored and un-scored data. There is a replication tool provided in the point-and-click interface, and an option to write code to simplify this process. There is no option to read in a key file to correspond with importing the data file. In addition, if one wants to change the score process the user must reimport and score the data again. The import step was somewhat difficult to implement. However, once the data was imported and scored, the remaining psychometric analyses were relatively simple.

Classical Test Theory Analyses

The CTT analyses included item and test analysis, and test scaling. The classical item analysis includes options such as the item statistics, reliability analyses, and conditional standard error of measurement (CSEM). The interface to generate this output is relatively easy to navigate. The output can be saved as a text file and includes all of the item statistics, test statistics, and reliability analysis. In the output each of the multiple choice (MC) and constructed response (CR) item options are provided which include the item difficulty, standard deviation, and two types of item correlations. In addition, there are five different methods of calculating reliability provided: Guttman's Lambda, Cronbach's Alpha, Feldt-Gilmer, Feldt-Brennan, and Raju's Beta.

Item analysis procedures also provide decision consistency and accuracy estimates: Huynh's Raw Agreement, Huynh's Kappa, KR-21, Beta-binomial alpha, and Beta-binomial beta. However, details on the method of calculation for some procedures are not clear. For example, the output has the Item-Total Pearson and Correlation Polyserial headers, but it is unclear if either option includes the current item in its correlation calculation. Another example is the CSEM, which is an option, but it is unknown which method is being used to calculate the results. The classical item analyses provided in jMetrik are quite comprehensive, useful to all levels of users both in research and applied settings. Documentation on procedures would be helpful but overall the interface is easy to use and output is complete.

The test scaling options are also easy to use and include many options. The user can quickly convert data to sum, percentile rank, Kelley True, and normalized scores. Users can specify constraints on the minimum, maximum, and precision points, as well as converting to a custom linear transformation. The program allows for renaming the new scored variables; however, only one transformation can occur in one run and once transformed there is no ability to rename variables. The test scaling features in jMetrik are useful, easy to use and understand. In summary, the CTT analyses provided by the program provide a simple point-and-click interface similar to SPSS and EXCEL, but provide the much needed psychometric analyses in one program that currently is unavailable for the psychometric specialist.

jMetrik allows for a more integrated system to conduct psychometric analyses for research and operational purposes without the cost associated with some other psychometric software programs.

Item Response Theory Analyses

The IRT analyses provided by jMetrik include calibration of the Rasch, Partial Credit and Rating Scale model. This program also has options for the calibration, item, and persons, which include the convergence criteria, scoring of missing data, fixing item parameter values, saving theta estimates and person fit statistics. The basic convergence options are available such as number of iterations, and convergence criteria, but the software is missing options such as estimation method, and theta estimate is produced (i.e., Expected a Priori, Maximum a Posteriori). The IRT statistics are saved in the data file, along with the original data sources, which can be exported to a text file.

IRT output includes the final item parameters, standard errors, WMS and UMS (which are assumed to stand for Weighted Mean Square and Unweighted Mean Squares), a score conversion table, and scale quality statistics. The output provided is helpful in evaluating the quality of the IRT analyses, but is missing some statistics that might be useful, such as item fit statistics. In addition, non-parametric item characteristic curves are provided in the program. Figure 1 shows an example item characteristic curve. The item characteristic curves can be printed or saved for each item or graph. There is no option to automatically save or print more than one graphic at a time.

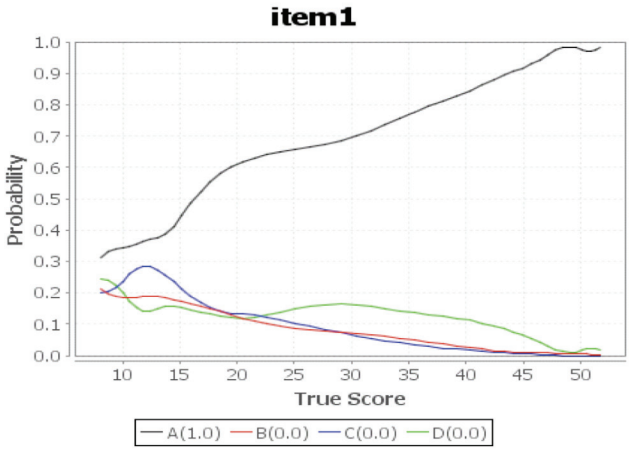


Figure 1: Sample item characteristic curve from jMetrik

In addition to the IRT calibration procedures, jMetrik can produce IRT equating results. The item parameters and thetas for the two sets of results need to be imported separately and in a particular format. Both item parameters and theta values along with weights are required to conduct the IRT equating and the program assumes that the Form Y values are the base form, and the Form X values will have the transformed theta and item parameters. The theta data file needs to have the theta value and weight, and the item parameter data file requires the item name and difficulty parameter.

The IRT equating output provides a robust z test, item summary statistics, and the equating coefficients for the mean/sigma, mean/mean, Haebara, and Stocking-Lord equating procedures. The IRT equating provides quick results and transformed both the item parameters and theta estimates. Test information graphics, or correlations of the item parameters, are not provided for before and after equating and the user needs to re-run each equating method if transformations are required from more than one equating procedure to calculate new theta values.

Differential Item Functioning

The non-parametric DIF analyses available are the Mantel Haentzel (MH) and non-parametric characteristic curves (CC) methods. The data for these analyses can be in a single file with an identifying variable indicating group membership. MH analyses provides two effect size measures and allows for matching either on observed or transformed Kelley scores. The CC analyses have options

to match on normal, true scores as well as several kernel regression options. The output is complete but difficult to interpret without the guidance of the FAQ's from the jMetrik website. There are many acronyms and unfamiliar terms that are not explained in the DIF output.

Documentation

The jMetrik software provides a quick start guide on the website that covers importing data, variable information, scoring item responses, and handling of missing data. There is also a ten-hour training session that is provided in January and August for a cost of $300. Unfortunately, a detailed user-manual was not available at the time of this review. The options provided in the software were fairly expansive, but full details on technical jargon are not outlined, which could be difficult to follow for new or occasional psychometric users. A detailed user-manual would provide much needed guidance and clarification and would enhance understanding of a useful program that users at all levels of expertise could appreciate.

The classical item analyses provided in jMetrik are quite comprehensive, useful to all levels of users both in research and applied settings.

Evaluative Conclusion

jMetrik provides psychometric software consumers with a program that offers a comprehensive assortment of analysis options. Current limitations in functionality of the program do detract from many of the features offered, but with revisions, the package could offer a user-friendly interface with complete psychometric analyses. For example, the CFA analyses offer several of the commonly used fit statistics, but some of the estimation procedures and options are under construction with options greyed out. Once all functionality is available a more thorough evaluation would be possible and the program should be very useful for the novice and expert psychometricians.

The lack of a user-manual makes it difficult to conduct analyses beyond data importing and scoring and interpret output. Additionally, a user-manual would allow for users to trouble-shoot some of the simple software errors as opposed to emailing the software developer for the solutions.

References

Drasgow, F., Luecht, R. M., & Bennett, R. E. (2006). Technology and testing. In R. L. Brennan (Ed.), *Educational measurement* 4th ed. (pp. 471-515). Westport, CT: Praeger Publishers.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading MA: Addison-Welsley Publishing Company

Notes in Brief

AUTHORS:

Carrie L. Zelna, Ph.D.
North Carolina State University

Stephany Brett Dunstan, Ph.D. Candidate
North Carolina State University

CORRESPONDENCE

Email
clzelna@ncsu.edu

AUTHORS' NOTE

The authors would like to thank the conference coodinators who provided helpful supplemental information about their conferences.

Finding the Right Fit: Choosing an Assessment Conference

There was a time when it could be a challenge to gain the necessary training and skills to conduct quality learning assessment. Fortunately, this is no longer the case as formal educational programs are now available through some graduate schools and, in addition, there are many outstanding assessment conferences. Faculty, staff and administrators sometimes struggle when researching the options for continuing education. The following pages provide the choices to consider when selecting an event as well as an abbreviated list of regional and national conferences.

Role of the Attendee

Attendees at assessment conferences tend to vary greatly in terms of their roles and levels of responsibility. As such, faculty, staff, and administrators often attend assessment conferences for different reasons. While faculty are focused on learning more about academic course, degree program and general education assessment, staff in the co-curricular and academic support programs are grappling with how to assess the co-curricular and support activities they deliver. Administrators and assessment professionals are focused on how best to organize a process and provide education and support to those in the degree programs or co-curricular units.

Scope of the Conference

The scope of conferences can also vary significantly, so when choosing a conference, it is important to identify the target audience, content areas and important logistics, as they will be key elements in the final decision.

Audience. Many conferences, but certainly not all, have a target audience based on the role of the attendees. Many conferences are general enough to provide valuable information for anyone engaging in the assessment process but often lean toward either an academic or co-curricular focus.

Content. One method of organizing content is to use "tracks." An assessment conference might have tracks associated with level of experience such as novice, intermediate and advanced. They might also have tracks associated with the area to be assessed such as Academic Programs, Student Affairs/Co-curricular programs, Academic Support and General Education. At times, tracks are associated with the type of institution, such as two- or four-year institutions. Some larger organizations have conferences where there are a variety of higher education topics and "assessment" itself is a track. In addition to tracks, most conferences offer pre-conference sessions and longer workshops that teach specific skills and/or allow you to work on a project or problem.

Attendees at assessment conferences tend to vary greatly in terms of their roles and levels of responsibility.

For guidance on what types of sessions you might want to attend, look to the various professional standards for assessment. The principles of assessment in Banta, Lund, Black & Oblander's (1996) book *Assessment in Practice: Putting Principles to Work on College Campuses* and the ACPA/NASPA professional standards for assessment (NASPA, 2012) may help guide your decisions.

Logistics. For budgetary reasons you might need to consider the logistics when choosing a conference. Location, size, length, sponsoring organization and inclusion of meals will impact the cost of registration and travel. The larger national conferences offer access to some of the best experts in the field but smaller conferences, even one-day drive-ins, could offer time with other practitioners and provide very useful support and personalized guidance.

In an attempt to demonstrate the variety of options, we have highlighted a number of conferences for your consideration. This list is arranged alphabetically and is by no means comprehensive. There are other high quality conferences we were unable to describe here.

American College Personnel Association (ACPA) Student Affairs Assessment Institute:
www.myacpa.org/pd/assessment/

The ACPA Student Affairs Assessment Institute is a national conference that meets in the spring and is geared toward student affairs professionals but inclusive of administrators, faculty and graduate students. With roughly 100 participants, small-group sessions are highlighted as a facet allowing for the development of individual assessment plans. The conference offers training and development in assessment at the beginning and intermediate levels over two and a half days. Tracks are offered for assessment at the department, program and the divisional level.

Association for Institutional Research (AIR) Annual Forum:
www.airweb.org/EducationAndEvents/AnnualConference/Pages/default.aspx

This national conference which meets in the spring touts itself on its website as being "the world's largest gathering of higher education professionals working in institutional research, assessment, planning and related postsecondary education fields." At the annual meeting over 400 presentations are offered in varying tracks, including at the 2012 meeting: Assessment: Accountability; Institutional Effectiveness, and Accreditation; Collaboration: Communicating Inside and Outside the Institution; Resources: Faculty, Finance and Facilities; Students: Enrollment and Experience; Technology: Data Management, Warehousing, and Internet and Computers.

Association for the Assessment of Learning in Higher Education (AALHE) Annual Conference:
www.aalhe.org

A new national conference in its second year, the AALHE is focused purely on student learning assessment in higher education. The quickly growing conference (130 attendees in its first year, but over 200 early registrants at the time of publication) is "open to all who are engaged in learning assessment—the faculty, administrators, Student Affairs

staff, institutional researchers and others in higher education who collect, analyze and use assessment data to improve student learning." The spring conference lasts roughly two and a half days, and the 2012 conference themes are Skills Necessary for Assessment Practice, Knowledge Necessary for Assessment Practice, Art of Assessment Practice, with strands at the beginning, intermediate and advanced level in: Teaching and Learning Theory; Curriculum Mapping and Development, Measurement of Learning, Data Analysis and Presentation, Communication/Documentation of Results. Special topics other than those listed here may also be included.

Indiana University-Purdue University Indianapolis Assessment Institute: www.planning.iupui.edu/institute

Meeting annually in the fall, this national conference indicates on its website that it is open to "faculty, student affairs professionals and administrators who have an interest in or responsibility for assessment" and generally has between 950 and 1,000 attendees. The two and a half day institute is held annually in the fall in Indianapolis, and tracks for the 2012 include Capstone Experiences, Civic Engagement, ePortfolios, Faculty Development, First-Year Experience, and Student Development.

International Association for Education Assessment Conference: www.iaea.info/

This international conference which meets annually in the fall has a mission to "help advance, through professional interchange, the science and practice of educational assessment by organizations around the world." The association's annual conference has been held in numerous locations around the globe including, in recent years, the Philippines, Thailand, Australia and the United Kingdom. Assessment topics included at the annual conference are broad, and range from literacy assessment in the classroom to national testing systems. Varying levels of education from elementary to higher education are included.

NASPA: Student Affairs Administrators in Higher Education Assessment & Persistence Conference: www.naspa.org/programs/apc/default.cfm

The NASPA Assessment & Persistence Conference is a national conference held annually in the spring with a mission to "promote student learning and success by strengthening assessment, quality improvement and intentional persistence and retention programs." The two and a half day conference generally hosts around 400 attendees and conference tracks are offered at the beginning, intermediate and advanced levels to accommodate attendees of varying backgrounds and levels of assessment experience, from faculty and student affairs professionals to assessment practitioners.

Many conferences are general enough to provide valuable information for anyone engaging in the assessment process but often lean toward either an academic or curricular focus.

New England Educational Assessment Network: www.neean.org/

This regional organization hosts both a fall forum and a summer institute and states on its website that its mission is "to promote quality assessment of student learning and development, and thus to enhance the effectiveness of institutions of higher education." Membership is open to faculty, staff, and administrators and the fall forum generally has roughly 150 attendees. The topics of presentations at the fall forum range from assessing student affairs and general education to graduate programs. The summer institute offers informational sessions as well a venue for working teams to craft assessment plans with the assistance of an NEAN consultant. The New England Educational Assessment Network also maintains a publication, the Journal of Assessment and Institutional Effectiveness.

Texas A&M Assessment Conference: www.assessment.tamu.edu/conference/

The Texas A&M Assessment Conference is a national conference meeting annually in the spring with a focus on student learning in higher education. The topics covered range from course-level assessment to general education and special interests. As such, the variety of topics is attractive to a wide audience in higher education from graduate students

to faculty and administrators, drawing roughly 600 participants. The two and a half day conference in 2012 offers the following tracks for concurrent sessions: Assessing Academic Programs & Projects; Faculty Assessing Student Learning at the Course Level; Strategic Planning and Assessment in Accreditation; Assessing General Education; Special Interest Topics: Assessing Student Affairs & Student Development; and Assessing Distance Learning.

Virginia Assessment Group Annual Conference: www.virginiaassessment.org/conferences_and_workshops.html

A regional conference now in its 25th year, the Virginia Assessment Group Annual Conference is held in the fall and is host to roughly 125-175 assessment practitioners, faculty and administrators with interests in assessment. The size of the conference offers ample opportunity for networking with practitioners. The 2012 meeting will feature sessions on best practices, research on assessment, and new and creative assessment practices.

Washington State Board for Community and Technical Colleges Assessment Teaching and Learning Conference: www.sbctc.ctc.edu/college/_e-assesspacnwteachinglearningconf.aspx

Some larger organizations have conferences where there are a variety of higher education topics and "assessment" itself is a track.

This regional conference is held annually in the spring and is geared toward community college educators. The conference generally has an audience of faculty from the community college system in Washington (as well as some from surrounding states) and the 2012 conference is expected to have roughly 400 attendees. Conference sessions are interactive and workshop-like with presentations by faculty for faculty to aid improving assessment practices in the classroom and beyond. Though there are no specifically designated themes or strands, topics at the 2012 meeting include writing course outcomes, aligning course outcomes with institutional learning outcomes, assessment practices in inverted classrooms, and electronic assessment management systems.

Western New York Assessment Consortium for Student Affairs and Enrollment Management: www.rit.edu/studentaffairs/assessment/consortium.php

A regional conference in its first year, the Western New York Assessment Consortium for Student Affairs and Enrollment Management serves student affairs and enrollment management professionals in higher education. This spring conference hosts roughly 100-150 participants and the website notes that "the consortium is designed to connect regional student affairs and enrollment management professionals to share and discuss assessment practices and knowledge." The day-long meeting include sessions highlighting best practices and Applied Assessment Sessions which include such topics as Residence life, Financial Aid, Career Services, Enrollment Services, Campus Life, and Orientation & Transition Programs.

Regional Accrediting Institutions' Annual Meetings

The regional accrediting agencies support assessment through conference sessions during the annual meetings, and in some cases, special assessment workshops throughout the year.

- New England Association for Schools and Colleges: http://www.neasc.org/
- Southern Association of Colleges and Schools: www.sacs.org/
- The Higher Learning Commission of the North Central Association of Colleges and Schools: www.ncahlc.org/
- The Middle States Commission on Higher Education: www.msche.org/
- Western Association of Schools and Colleges: http://wascarc.org/

As you search for the right conference, keep in mind that this list is abbreviated and in addition to these types of conferences, academic disciplines and co-curricular and academic support programs often have organizations that host national conferences. These conferences often include an assessment component. More information regarding conferences can be found online or through your professional association.

References

Banta, T.W., Lund, J.P., Black, K.E., & Oblander, F.W. (1996). *Assessment in practice: Putting principles to work on college campuses.* San Francisco, CA: Jossey-Bass.

National Association of Student Personnel Administrators. (2012). Assessment, evaluation and research. NASPA. Retrieved from http://naspa.org/programs/profdev/assessment.cfm

RUMINATE: INTEGRATING THE ARTS AND ASSESSMENT

Photographer: Katie Basbagill
Indianapolis, IN
www.bohemianredimages.com

THE BLIND MEN AND THE ELEPHANT

A Hindoo Fable

By John Godfrey Saxe (1816–1887)
19th Century American Poet

It was six men of Indostan
To learning much inclined,
Who went to see the Elephant
(Though all of them were blind),
That each by observation
Might satisfy his mind.

The First approached the Elephant,
And happening to fall
Against his broad and sturdy side,
At once began to bawl:
"God bless me!—but the Elephant
Is very like a wall!"

The Second, feeling of the tusk,
Cried:" Ho!—what have we here
So very round and smooth and sharp?
To me 't is mighty clear
This wonder of an Elephant
Is very like a spear!"

The Third approached the animal,
And happening to take
The squirming trunk within his hands,
Thus boldly up and spake:
"I see," quoth he, "the Elephant
Is very like a snake!"

The Fourth reached out his eager hand,
And felt about the knee.
"What most this wondrous beast is like
Is mighty plain," quoth he;
"'T is clear enough the Elephant
Is very like a tree!"

The Fifth, who chanced to touch the ear,
Said: "E'en the blindest man
Can tell what this resembles most;
Deny the fact who can,
This marvel of an Elephant
Is very like a fan!"

The Sixth no sooner had begun
About the beast to grope,
Than, seizing on the swinging tail
That fell within his scope,
"I see," quoth he, "the Elephant
Is very like a rope!"

And so these men of Indostan
Disputed loud and long,
Each in his own opinion
Exceeding stiff and strong,
Though each was partly in the right,
And all were in the wrong…

GUIDELINES FOR SUBMISSION

All manuscripts submitted to *Research & Practice in Assessment* may be related to various higher education assessment themes, and should adopt either an assessment measurement or an assessment policy/foundations framework:

Assessment Measurement:
a) instrument design, b) validity and reliability, c) advanced quantitative design, d) advanced qualitative design

Assessment Policy/Foundations:
a) accreditation, b) best practices, c) social and cultural context, d) historical and philosophical context, e) political and economic context

Article Submissions:

Articles for *Research & Practice in Assessment* should be research-based and include concrete examples of practice and results in higher education assessment. The readers of *Research & Practice in Assessment* are associated with myriad institutional types and have an understanding of basic student learning and assessment practices. Articles for publication will be selected based on their degree of relevance to the journal's mission, compliance with submission criteria, quality of research methods and procedures, and logic of research findings and conclusions. Approximately fifty percent of submissions are accepted for publication.

Review Submissions:

Reviews (book, media, or software) are significant scholarly contributions to the education literature that evaluate publications in the field. Persons submitting reviews have the responsibility to summarize authors' works in a just and accurate manner. A quality review includes both description and analysis. The description should include a summary of the main argument or purpose and overview of its content, methodology, and theoretical perspective. The analysis of the book should consider how it contrasts to other works in the field and include a discussion of its strengths, weaknesses and implications. Judgments of the work are permitted, but personal attacks or distortions are not acceptable as the purpose of the review is to foster scholarly dialogue amongst members of the assessment community. The *RPA* Editor reserves the right to edit reviews received for publication and to reject or return for revision those that do not adhere to the submission guidelines.

Special Features:

Each issue of *Research & Practice in Assessment* highlights the work of a noted researcher or assessment professional in a manner that aims to extend the scholarly dialogue amongst members of the assessment community. Special Features are invited by the Board of Editors and often address the latest work of the author.

Notes in Brief:

Notes in Brief offer practitioner related content such as commentaries, reports, or professional topics associated with higher education assessment. Submissions should be of interest to the readership of the journal and are permitted to possess an applied focus. The *RPA* Editor reserves the right to edit manuscripts received for publication and to reject or return for revision those that do not adhere to the submission guidelines.

Ruminate:

Ruminate concludes each issue of *Research & Practice in Assessment* and aims to present matters related to educational assessment through artistic medium such as photography, poetry, art, and historiography, among others. Items are encouraged to display interpretive and symbolic properties. Contributions to Ruminate may be submitted electronically as either a Word document or jpg file.

Manuscript format requirements available at: www.RPAjournal.com

RPA BOOKS AVAILABLE

Astin, A. W. (2012). *Assessment for excellence: The philosophy and practice of assessment and evaluation in higher education (2nd ed.).* Lanham, MD: Rowman & Littlefield. Pp. 374. $55.00 (paper).

Balzer, W. K. (2010). *Lean higher education: Increasing the value and performance of university processes.* New York, NY: Productivity Press. pp. 312. $51.95 (paper).

Bresciani, M. J., Gardner, M. M., & Hickmott, J. (2010). *Demonstrating student success: A practical guide to outcomes-based assessment of learning and development in student affairs.* Sterling, VA: Stylus. pp. 224. $27.50 (paper).

Butt, G. (2010). *Making assessment matter.* New York, NY: Continuum International. pp.160. $27.95 (paper).

Cambridge, D. (2010). *Eportfolios for lifelong learning and assessment.* Hoboken, NJ: Wiley, John & Sons.pp. 288. $38.00 (hardcover).

Carey, K. & Schneider, M. (Eds.). (2010). *Accountability in American higher education.* New York, NY: Palgrave Macmillan. pp. 355. $95.00 (hardcover).

Christensen, C. & Eyring, H. (2011). *The innovative university: Changing the DNA of higher education from the inside out.* San Francisco, CA: Jossey Bass. pp. 512. $32.95 (hardcover).

Côté, J. E. & Allahar, A. L. (2011). *Lowering higher education: The rise of corporate universities and the fall of liberal education.* University of Toronto Press. pp. 256. $24.95 (paper).

Dunn, D. S., McCarthy, M. A., Baker, S. C., & Halonen, J. S. (2010). *Using quality benchmarks for assessing and developing undergraduate programs.* San Francisco, CA: Jossey Bass. pp. 384. $45.00 (hardcover).

Flateby, T. L. (Ed.). (2010). *Improving writing and thinking through assessment.* Charlotte, NC: Information Age. pp. 238. $45.99 (paper).

Hutchings, P., Huber, M. T., & Ciccone, A. (2011). *The scholarship of teaching and learning reconsidered: Institutional integration and impact.* San Francisco, CA: Jossey Bass. pp. 224. $30.00 (paper).

Joughin, G. (Ed.). (2009). *Assessment, learning and judgment in higher education*. New York, NY: Springer. pp. 445. $289.00 (hardcover).

Kramer, G. L. & Swing, R. L. (Eds.). (2010). *Higher education assessments: Leadership matters.* Lanham, MD:Rowman & Littlefield. pp. 288. $ 49.95 (hardcover).

Maki, P. L. (2010). *Assessing for learning: Building a sustainable commitment across the institution.* Sterling, VA: Stylus. pp. 356. $32.50 (paper).

Martin, R. (2011). *Under new management: Universities, administrative labor, and the professional turn.* Philadelphia, PA: Temple University Press. pp. 253. $69.50 (cloth).

Noyce, P. E. & Hickey, D. T. (Eds). (2011). *New frontiers in formative assessment*. Cambridge, MA: Harvard Education Press. pp. 260. $29.95 (paper).

Rothman, S., Kelly-Woessner, A., & Woessner, M. (2010). *The still divided academy: How competing visions of power, politics, and diversity complicate the mission of higher education.* Lanham, MD: Rowman & Littlefield. pp.296. $39.95 (cloth).

Schloss, P. J. & Cragg, K. M. (2012). *Organization and administration in higher education.* New York, NY: Routledge. pp. 350. $160.00 (hardcover).

Secolsky, C. & Denison, B. D. (Eds). (2011). *Handbook on measurement, assessment, and evaluation in higher education.* New York, NY: Routledge. pp. 704. $119.95 (paper).

Shavelson, R. J. (2009). *Measuring college learning responsibly: Accountability in a new era.* Palo Alto, CA: Stanford University Press. pp. 256. $21.95 (paper).

Stone, T. & Coussons-Read, M. (2011). *Leading from the middle: A case-study approach to academic leadership for associate deans.* Lanham, MD: Rowman & Littlefield. pp. 208. $40.00 (cloth).

Tehmina, N. B. (2010). *Conducting research in educational contexts.* New York, NY: Continuum International. pp. 256. $34.95 (paper).