

Review of Practical Assessment, Research & Evaluation, Volume 10

Amy Gullickson

Western Michigan University

Practical Assessment, Research & Evaluation is a peer-reviewed online journal that publishes articles individually throughout the year, which is edited by Lawrence Rudner and William Schafer. The chronological listing of contributions for a particular year constitutes a volume. The articles reviewed here are some of the recent entries from this year (Volume 10, 2005), and can be found at <http://pareonline.net>.

In “Enhancing Validity in Phonological Awareness Assessment through Computer Supported Testing” (Volume 10 (18), 2005), Jerrell C. Cassady, Lawrence L. Smith & Linda K Huber, discuss how their Standardized Assessment of Phonological Awareness (SAPA) addresses threats to validity that other testing methods do not. The problems with the assessment tools currently in use are twofold, according to Cassady, Smith and Huber. First, they are too broad – assessing domains, rather than discrete abilities – and therefore not achieving an accurate measure of a child’s phonological awareness. Second, oral presentation of the prompts in these instruments creates unavoidable variation from person to

person. The SAPA addresses both of these problems. Accurate measure is created by assessing awareness in three discrete sets of tasks: phonemic position, phonological awareness, and linguistic unit. Presentation is standardized through the use of pre-recorded, digitally spaced stimuli which are given to students via computer. Testing for validity revealed that the SAPA had a strong positive correlation for total scores compared with student reading outcomes. Reliability testing showed strong internal consistency. The added benefit of new format is that training a teacher to appropriately administer the computer test takes 15 minutes, whereas training to give the test orally takes 2 hours plus follow-up work.

Brian Noonan & C. Randy Duncan explored “Peer and Self-Assessment in High Schools” (Volume 10 (17), 2005). Their literature search indicated that both peer and self-assessment are important contributors to excellent formative evaluation of students. Noonan and Duncan then studied data gathered from a survey of 118 high school teachers’ assessment practices to determine whether, and to what extent, those practices are actually being used in the classroom. According to their analysis of the qualitative data, 76% of high school teachers use peer or self-assessment “a little” or “somewhat” for formative evaluation (p.4). Teachers discussed using peer assessment primarily to facilitate group work and activities, while self-assessment was aimed at helping students reflect on their own performance. Noonan and Duncan saw four areas for possible future research in this area: 1) the nature and extent of student involvement in peer or self-assessment, 2) the quality of processes used for those assessments, 3) the ways in which these practices can be used to improve instruction and 4) the role of the teacher when these assessments are used.

Heping Deng conducted a power analysis on 80 quantitative, successfully defended dissertations from universities in Tennessee in his article, “Does It Matter If Non-Powerful Significance Tests Are Used in Dissertation Research?” (Volume 10 (16), 2005). Power is the calculated ability for a statistical test to discern effects within the sample. Deng’s research indicated that the power levels in all 80 dissertations had sufficient strength to detect large effects at Cohen’s .80 power criterion. However, they did not have enough power to accurately detect medium and small effect size. In addition, the sample sizes used were 2.5 the average optimal size. Therefore, Deng recommends training students in power analysis, so that they can use it to determine *a priori* their optimal sampling frame for measuring their chosen effect size. He also believes that effect size and statistical power need to be calculated and reported for all statistical tests.

Richard P. Phelps attacks the “Greene Method” in “A Review of Greene (2002) High School Graduation Rates in the United States” (Volume 10 (15), 2005). Phelps declares that Greene fails to understand 1) the reality of data collected on high school populations, 2) the difference between rate and ratio, and 3) the appropriate generalizability of findings. First, according to Phelps, data collection agencies like the Organization for Economic Cooperation and Development (OECD) often have inaccurate data due to variations in the way schools collect and report their numbers. In addition, the decision on how to measure dropouts is fuzzy at best, so those statistics are unlikely to be comparable from school to school. Phelps claims that Greene does not take into account the lack of accuracy and presence of noncomparable measures in his creation of graduation rates, nor does he place any limitations on his findings due to those considerations. Second, Phelps states that Greene fails to understand the difference between rates and

ratios. Rates require “all the particular individuals counted in the numerator also be present in the denominator count” (p.12). A ratio can be any one number divided by another number. In the case of high school graduation, if a student moves into the district after the year’s official head count has been conducted, he or she must be added to the total in order to accurately calculate the school’s graduation rate (# of graduates/ # of total students in grade 12). In contrast, the completion ratio can be calculated using the number of persons of their same age living in the area, regardless of student mobility in and out of a particular school (# of graduates/ # of 18 year olds within the district). Both systems have flaws. However, Greene’s use of the graduation rate fails to sufficiently take into account that student migration and fluctuations in population and fertility can impact a district’s number of graduates (numerator), but those figures may not have been included in the official head count of students in that grade at that particular school (denominator). Thus, his calculations are not as valid as he markets them to be. Finally, Greene claims that his measure can be used to judge performance across jurisdictions because of its consistency. Phelps argues that his failure to adequately address demographic trends within his calculations creates test bias. Phelps drives his point home by analyzing graduation rates of two fictional towns, demonstrating that the Greene method miscalculates by 6.5 percentage points in one, and 12 percentage points in another.

References

Cassady, J. C., Smith, L. L., & Huber, L. K. (2005). Enhancing validity in phonological awareness assessment through computer-supported testing.

Practical Assessment Research & Evaluation, 10(18). Available at

<http://pareonline.net/getvn.asp?v=10&n=18>

Deng, H. (2005). Does it matter if non-powerful significance tests are used in dissertation research. *Practical Assessment Research & Evaluation*, 10(16).

Available at <http://pareonline.net/getvn.asp?v=10&n=16>

Noonan, B., & Duncan, C. R. (2005). Peer and self-assessment in high schools.

Practical Assessment Research & Evaluation, 10(17). Available at

<http://pareonline.net/getvn.asp?v=10&n=17>

Phelps, R. P. (2005). A Review of high school graduation rates in the United States. *Practical Assessment Research & Evaluation*, 10(15). Available at

<http://pareonline.net/getvn.asp?v=10&n=15>