

SOCIAL WORK WRITTEN COMPREHENSIVE EXAM

ASSESSMENT PRACTICE & RECOMMENDATIONS

Becca Marsh, Kelli Samonte, Alan Socha, & Matthew Swain

TABLE OF CONTENTS

Background of the Project	3
Results.....	4
Content Alignment	4
Standard Setting.....	4
Item Analysis.....	5
Aggregate Data on the SWWC Exam.....	5
Problematic Items.....	5
Interviews with B.S.W. Students (Qualitative)	7
Recommendations.....	7
Program Improvement.....	7
Assessment Design Improvement	7
Reliability and Validity	7
Content Alignment	8
Standard Setting	8
Item Analysis.....	8
Qualitative Data.....	9
Proposed Strategy	9
References	10

BACKGROUND OF THE PROJECT

The purpose of this program assessment project was to facilitate the piloting and instrument validation process for the James Madison University (JMU) B.S.W. program Social Work Written Comprehensive (SWWC) exam. The SWWC exam is a scenario-based, multiple-choice instrument developed by JMU undergraduate Social Work faculty and administered via Qualtrics survey software. The 60-item SWWC exam was developed to evaluate student competency on a number of undergraduate student learning objectives developed for the Social Work program. These objectives correspond to competencies specified in 2008 by the Council on Social Work Education (CSWE), which serves as the accreditation agency for the JMU Social Work undergraduate program. JMU faculty specifically created the SWWC exam items to align with and evaluate undergraduate Social Work student competencies in regards to 10 general requirements, or Educational Policy and Accreditation Standards (EPAS), as required by CSWE for practicing Social Workers in the United States. According to the EPAS competencies outlined by CSWE, Social Work students are considered to be competent once they have demonstrated knowledge and/or proficiency in the following domains (CSWE, 2008):

- 2.1.1 Identify as a professional social worker and conduct oneself accordingly
- 2.1.2 Apply social work ethical principles to guide professional practice
- 2.1.3 Apply critical thinking to inform and communicate professional judgments
- 2.1.4 Engage diversity and difference in practice
- 2.1.5 Advance human rights and social and economic justice
- 2.1.6 Engage in research-informed practice and practice informed research
- 2.1.7 Apply knowledge of human behavior and the social environment
- 2.1.8 Engage in policy practice to advance social and economic well-being and to deliver effective social work services
- 2.1.9 Respond to contexts that shape practice
- 2.1.10 Engage, assess, intervene, and evaluate with individuals, families, groups, organizations, and communities

JMU seniors earning a bachelor's degree in Social Work (B.S.W.) must demonstrate proficiency in these domains by passing the SWWC exam and an oral comprehensive exam. Specifically, they must pass the SWWC in order to be eligible to complete the Social Work oral comprehensive exam and then graduate. The SWWC exam is criterion-referenced. That is, JMU Social Work faculty will establish a criterion for B.S.W. student performance. All B.S.W. students will be expected to meet this criterion before moving on to complete the Social Work oral comprehensive exam. In order to make high-stakes decisions regarding B.S.W. student performance on the SWWC exam, the Social Work program collaborated with JMU Center for Assessment and Research Studies (CARS) to engage in SWWC exam pilot testing and an extensive instrument validation process that included: content alignment, standard-setting, item analysis, qualitative interviews with B.S.W. students post SWWC exam pilot testing, and compilation of SWWC exam reliability and validity evidence. *The Program Evaluation Standards* (Yarbrough, Shulha, Hopson, & Caruthers, 2011) and *The Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999) were both adhered to and used to help direct this process.

RESULTS

CONTENT ALIGNMENT

All items were written with the intention of measuring one or more of the EPAS competencies except items 20 and 22, which were written to assess competencies added by the Social Work faculty. Additionally, not all EPAS competencies are assessed equally by the SWWC exam. As this exam is one component of a larger assessment plan, it is not the purpose of the test to assess demonstration of every competency. However, the appropriate alignment of the items to competencies was evaluated via a content alignment. In essence, content alignment is a process by which raters map each item to the objective in which they believe it measures. Two different methods were used to conduct the content alignment; however, analyses of the data separately by method and when methods were combined led to somewhat undesirable results.

The item-objective mappings differed across content alignment methods, which was expected (Miller, Setzer, Sundre, & Zeng, 2007). What was troubling was that neither method produced the same mappings as specified by the Social Work faculty. For example, very few raters mapped any items to Objective 9 although multiple items were originally written to measure that objective. Finally, Cohen's Kappa (κ) estimates indicated that there was either "slight" or "poor" agreement amongst raters for most objectives. For more detail regarding the content alignment procedure and results, please refer to the Content Alignment report and the Test Manual (Marsh, Samonte, Socha, & Swain, 2011a, 2011b).

The lack of alignment between items and the appropriate objective(s) could pose several issues for the Social Work faculty. Ultimately, it indicates at face value that the items may not measure what they were intended to measure. This may be a threat to validity. However, it is important to note that the raters of the content alignment study were not subject matter experts and thus the content alignment should be replicated with independent subject matter experts.

STANDARD SETTING

After piloting the SWWC exam, but before viewing the results from the fall 2011 semester, the Social Work faculty completed a standard setting process in order to establish a performance criterion for the SWWC exam. Social Work faculty members were asked to categorize the 19 B.S.W. students who would be completing the pilot of the SWWC exam during the fall 2011 semester into various categories: students they believed would perform the best on the exam (the "High" group), students they believed would perform poorly on the exam (the "Low" group), and students that they considered to be "borderline" or straddling the line between low and high performance on the exam (the "Borderline" group). For more detail regarding the standard setting process, please refer to the SWWC exam Test Manual (Marsh et al., 2011b).

SWWC exam performance was very similar for all students. Means were computed for each group (i.e., "high," "low," and "borderline") via different levels of faculty agreement. In each case the means that were essentially the same among groups. Therefore, the results are nonsensical and are not helpful to Social Work faculty in classifying students because of the

minimal amount of variability in scores. In order to establish a meaningful SWWC exam criterion, Social Work faculty may consider conducting this process again with the spring 2012 B.S.W. sample or consider an alternate standard setting method.

ITEM ANALYSIS

Recall, a total of 19 students completed the SWWC pilot exam in the fall of 2011. A summary of aggregate data regarding student performance on the SWWC exam is provided below. Additionally, item analyses were conducted using the pilot sample. Examples of problematic items are also presented. For a detailed narrative of the item analysis process, refer to the SWWC exam Test Manual (Marsh et al., 2011b).

AGGREGATE DATA ON THE SWWC EXAM

Student mean total exam score was 33.34 out of a possible 60 points. This corresponds to a mean total percent correct score of 55.57%, indicating that, on average, students are responding correctly to just over half the items on the exam. Total scores ranged between 25 (41.67% correct) and 40 (66.67% correct). Notably there was little variability among the scores of the 19 Social Work students. Reliability was calculated for the total scores ($\alpha = .43$). In general, reliability estimates of .70 or larger indicate that total scores are interpretable and can be used for research or program evaluation purposes. Although this estimate is low, it should be noted that it was calculated based upon a very small, non-representative pilot sample ($N = 19$). Thus, we cannot draw definitive conclusions about the SWWC exam reliability evidence based only upon this administration—further piloting is needed.

PROBLEMATIC ITEMS

An initial review of item analyses for the SWWC exam was presented during the Social Work assessment committee meeting held on October 12th, 2011. These results revealed that there were several items that did not function well with the pilot sample as they appeared to be too difficult (i.e., none of the 19 students answered them correctly) or too easy (i.e., all 19 students answered them correctly). Although many items were problematic, a few examples of these items are presented below. Other problematic items may be identified from the item analyses listed in the Test Manual for the SWWC exam (Marsh et al., 2011b).

For example, items 10-15 below, students were asked to order the steps in the correct sequence for the scenario. No student was able to respond with the correct sequence of steps.

Some clients coming through the Lamb Center are young men age 18 – 30 who do not have a high school diploma or GED; most have completed some high school. Clients have told you this is a barrier to employment.

For items 10-15, order the steps you might take to address this concern of clients using a quantitative perspective.

- a. Collect data from agency participants
- b. Write a report on the impact of a high school diploma or GED on the employment outcomes of homeless men who come to the Lamb Center
- c. Search for information on the education and employment barriers
- d. Select methods to measure the numbers of clients coming to the agency who do not have high school diplomas or GEDs
- e. Develop a formal research question
- f. Analyze data collected

For item 55 below, all 19 students responded incorrectly with “C” instead of the correct response “B”. While good distractors help to improve the measurement of the correct skills rather than misconceptions, option B was an attractive option for all students. The Social Work faculty must decide if distractors that function in such a way should be removed.

55. In trying to formulate a specific research question to direct the community needs assessment, you decide the following is not helpful:

- a. Talk with members at professional social work organizations about your study
- b. Look in Wikipedia for information
- c. Use resources provided by diverse groups/organizations
- d. Speak with agency personnel who have expressed similar client concerns

For item 35 below, all 19 students correctly responded with “A” to the item. While we hope that there are some core concepts in which all students answer correctly, it is important to remember that items that may be too easy do not contribute to discrimination among students.

35. At Pleasantide Assisted Living Facility most activities for residents involve crafts, sewing, or religious activities. Female residents appear to find it easy to participate while male residents sit along the wall watching. You recognize this can hasten increased fragility and impairment. This has struck a curiosity in you about activities for male residents. What would be the BEST course of action?

- a. Speak to male residents individually to see what type of activities they would like to participate in at the facility
- b. Get the group facilitator to include the men in the sewing group
- c. Ask female residents to include the males in their activities
- d. Do nothing, it is not a core responsibility of your job description

Too many problematic items are a threat to test validity as the test may assess irrelevant outcomes and add noise to the total scores. These item issues should be addressed considering the high stakes nature of the test. Difficult or unfair items will inherently lower scores and may not assess the competency that it was intended to measure. Additionally, too many easy items can unintentionally inflate student scores.

INTERVIEWS WITH B.S.W. STUDENTS (QUALITATIVE)

Students were interviewed after completing the SWWC exam. Their responses gave insight as to the student views of exam administration, exam content, program preparation and student preparation. Overall students tended to be satisfied with the way the exam was administered, the content included on the exam, the preparation provided by the program, and their personal preparation. Several students also provided suggestions for future administrations such as providing scrap paper for students to take notes and providing more case studies as preparatory examples. The interviews with the students can help to inform Social Work faculty of student perceptions and issues regarding several aspects of the SWWC exam. Please refer to the Test Manual for a detailed synopsis of the interview themes (Marsh et al., 2011b).

RECOMMENDATIONS

PROGRAM IMPROVEMENT

Programmatic changes based on the results from the SWWC thus far are not recommended. Additional pilot testing, validation procedures and possible revisions should be made before any substantial changes are implemented. Interview data suggested that students would have preferred to meet earlier in the semester to discuss the SWWC exam. They also suggested that the specific EPAS competencies covered in each course could be more explicitly highlighted.

ASSESSMENT DESIGN IMPROVEMENT

The assessment design of the Social Work program includes many components to examine student learning and competency. As one component of the assessment design, weaknesses of the SWWC exam will weaken interpretations of program effectiveness. Improvement of the SWWC exam strengthens the overall assessment design of the Social Work program. Most important areas of improvement include reliability and validity concerns.

RELIABILITY AND VALIDITY

Coefficient alpha may not be the most appropriate estimate of reliability for the SWWC exam as it is competency-based and criterion-referenced. More specifically, coefficient alpha is more appropriate when an instrument is norm-referenced and there is a fair amount of variance in respondent scores. Oftentimes, there is not as much variability in respondent scores for competency-based, criterion-referenced instruments. Thus, it would be most appropriate to use an index of decision consistency when bolstering SWWC exam reliability evidence. Decision consistency index estimates such as the Subkoviak Index of Decision Consistency (Subkoviak, 1976) indicate whether respondents are reliably classified in regards to a competency-based instrument criterion (e.g., the pass/fail cut-score standard established by Social Work faculty) based upon a single test administration. Subkoviak's Index of Decision Consistency (Subkoviak,

1976) can and should be calculated based upon aggregate data from the fall 2011 and spring 2012 administrations.

CONTENT ALIGNMENT

Because the content alignment study was not conducted by subject matter experts, the content alignment should be replicated with subject matter experts. Additionally, it may be helpful to choose one method when conducting a content alignment. The objective-by-objective method has been shown to facilitate multidimensional item mappings and thus may be best suited for the SWWC exam (Miller et al., 2007).

STANDARD SETTING

Again, additional pilot studies should be conducted with larger samples of students. For instance, the standard setting procedure conducted in the fall of 2011 resulted in standards with little meaning due to the low variability in student scores ($N=19$). A larger sample collected in the spring semester of 2012 may have more variability to derive meaningful standards.

While the Social Work faculty members have ample knowledge regarding individual student ability, and thus are well suited for student-centered standard setting processes, other methods may be more appropriate and helpful in the standard setting process. Many of the methods ask faculty to imagine a “borderline” student. For example, the bookmark method orders items from easiest to most difficult and then asks faculty to identify the item at which a hypothetical “borderline” student would answer incorrectly. Additionally, the Nedelsky method can be used with multiple choice items. This method asks faculty to examine each item and estimate the number of response options a borderline participant would be able to rule out as incorrect. In the Jaeger method, faculty members are not asked to imagine a “borderline” participant. Instead, this method asks faculty to rate items with a “yes” or “no” as to whether or not *all* participants should be able to answer the item. This method shifts the focus away from a hypothetical group but may lead to impractically high standards. Ultimately, Social Work faculty should consider the many options for standard setting and choose the one in which they feel most comfortable. For a review of other standard setting methods see Nichols, Twing, Mueller, and O’Malley (2010) and Hambleton and Pitoniak (2006).

ITEM ANALYSIS

Items that asked students to arrange the steps of a process or procedure posed a difficulty with scoring. Specifically, credit for item groups 10-15 and 39-44 was assigned based on a “close” answer to the correct order; however, no student received full credit for these item groups. The Social Work faculty should revisit these items. If the concepts assessed by these items are deemed important, perhaps a different format would allow for students who have a high understanding of the concept to receive full credit. Additionally, any items that were too easy or too difficult should be revisited. These items can be retained for construct coverage, revised to make them more or less difficult, or omitted.

QUALITATIVE DATA

The following changes are suggested based on student concerns of the computer-based SWWC exam:

- (1) Unfortunately, Qualtrics will not allow the test taker to highlight and underline parts of the scenarios and will not allow for the scenario to be “frozen” at the top of the page to eliminate the need for scrolling. Despite this, use of Qualtrics still allows for an easier, more seamless implementation of the test. Non-technological methods can be used, instead, to work around the software limitations and accommodate the students’ concerns. Students could be supplied a piece of scrap paper with the scenarios allowing for highlighting and note taking while answering items on the computer.
- (2) Progress bars may be implemented on each page to give students an indication of their progress through the exam.
- (3) Displaying the student’s grade immediately upon completion can be implemented in Qualtrics.

If the above suggestions are implemented, careful attention must be made regarding test security. Specifically, all scrap paper should be collected before the student leaves the room so the scenarios and test questions are not leaked. Displaying student grades after completing the exam should follow any changes made to the scoring procedure. The way “ordering” items are currently being scored (i.e., partial credit for close responses) cannot be automatically scored by Qualtrics. Thus the scoring of the “ordering” items should be discussed if faculty members would like to present scores automatically.

PROPOSED STRATEGY

Based on the results presented above, several steps should be taken by Social Work faculty before student-level or program-level decisions are made. Much of the work to be done must wait until the next pilot test administration (spring 2012). In the near future, Social Work faculty should review results of the content alignment and item analysis from fall 2011. Faculty should note any major issues and consider how to address such issues in the future.

As previously mentioned, Social Work faculty should pilot test this instrument with the spring 2012 cohort. Qualitative responses from the students should be taken into account and incorporated where possible (i.e., providing scratch paper). Faculty should also choose the standard setting method that would best address their needs and conduct that method for the spring 2012 sample. After pilot testing, item analyses should be conducted on all 60 items to determine if any items that seem to be problematic. These items should be revised or omitted. Additionally, content alignment using independent content experts should be considered.

REFERENCES

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Berk, R. A. (1976). Determination of optional [sic] cutting scores in criterion-referenced measurement. *Journal of Experimental Education*, 45(2), 4-9.
- Black, P. N., & Whelley, J. (1999). The social work licensure exam: examining the exam through the lens of CSWE curriculum policy, *Arête*, 23, 66–76.
- Council on Social Work Education (CSWE). (2008). Educational policy and accreditation standards. Retrieved from <http://www.cswe.org/File.aspx?id=13780>.
- DeLong-Hamilton, T., Buchan, V. V., Hull, G. H., Christenson, B., Gerritsen-McKane, R., Rodenhiser, R. W., & Smith, M. L. (2011) Responding to the 2008 EPAS: Baccalaureate education assessment direct and indirect measurement. *Journal of Baccalaureate Social Work* 16(1), 17-31.
- Hambleton, R. K. & Pitoniak, M. J. (2006). Setting performance standards. In R. L. Brennan (Ed.), *Educational measurement*. (pp. 433-470). Westport, CT: Praeger Publishers.
- Marsh, B., Samonte, K., Socha, A., & Swain, M. (2011a). *Social Work Written Comprehensive exam content alignment*. Harrisonburg, VA: James Madison University, Center for Assessment and Research Studies.
- Marsh, B., Samonte, K., Socha, A., & Swain, M. (2011b). *Social Work Written Comprehensive exam test manual*. Harrisonburg, VA: James Madison University, Center for Assessment and Research Studies.

- Miller, B. J., Setzer, C., Sundre, D. L., & Zeng, X. (2007). *Content validity: A comparison of two methods*. Paper presented to the National Council on Measurement in Education, Chicago, IL.
- Nichols, P., Twing, J., Mueller, C.D., & O'Malley, K. (2010). Standard-setting methods as measurement processes. *Educational Measurement: Issues and Practice*, 29, (1), 14-24.
- Subkoviak, M. J. (1976). Estimating reliability from a single administration of a criterion-referenced test. *Journal of Educational Measurement*, 13, 265-276.
- Yarbrough, D. B., Shulha, L. M., Hopson, R. K., and Caruthers, F. A. (2011). *The program evaluation standards: A guide for evaluators and evaluation users (3rd ed.)*. Thousand Oaks, CA: Sage.
- Zieky, M. J., & Livingston, S. A. (1977). *Manual for setting standards on the Basic Skills Assessment Tests*. Princeton, NJ Educational Testing Service.