An Integrative Framework for Developing Comprehensive Examinations to Measure Student Social Work Competencies

Todd M. Jensen and Kimberly J. Strom-Gottfried

ABSTRACT
Social work and other professions are endeavoring to promote student proficiency with respect to core competencies of the discipline, and the attainment of these competencies must be demonstrated objectively in some way. This article provides an illustrative guide for social work educators seeking to develop comprehensive examinations to assess students’ acquisition of professional competencies, particularly at the master’s level and across generalist and specialization levels of practice. An integrative framework is presented, highlighting 5 central stages of examination development: planning, creating and implementing, evaluating, revising, and perpetuating. These stages are illustrated with a case example, followed by key conclusions, methodological considerations, and other implications for moving forward on efforts to effectively develop examinations to measure student social work competencies.

ARTICLE HISTORY
Accepted: August 2017

The assessment of student learning and competence is a prominent topic in contemporary higher education (Astin & Antonio, 2012; Zlatkin-Troitschanskaia, Shavelson, & Kuhn, 2015). This is driven, in part, by accrediting and funding bodies that have demanded advances in student outcome assessment and institutional accountability (Astin & Antonio, 2012). Moreover, the professional disciplines are focused on student development and readiness with respect to core professional competencies, whose attainment must be demonstrated objectively in some way (Banta, Jones, & Black, 2009).

Following debate earlier this century about social work accreditation standards and student evaluation processes (e.g., Gambrill, 2001; Stoesz & Karger, 2009), the profession has largely moved on, embracing accreditation processes and implementing competency-focused education and experiences. Curricula are being designed or adapted to promote students’ acquisition of specific professional competencies to best prepare them for generalist and specialized levels of social work practice. Social work competencies are described in the Educational Policy and Accreditation Standards (EPAS) by the Council on Social Work Education’s (CSWE, 2015) Commission on Accreditation. The EPAS also mandates that social work programs “assess students’ demonstration of the social work competencies through the use of multi-dimensional assessment methods” (CSWE, 2015, p. 19). Common forms of student assessment in social work include field evaluations, embedded measures in course assignments, and cumulative capstone or portfolio projects. Local, or curriculum-specific, comprehensive assessment instruments are also promising yet underused tools for measuring and comparing student competency outcomes (Drisko, 2014; Patterson, 2006; Zlatkin-Troitschanskaia et al., 2015). Thus, social work educators should be equipped with the knowledge and skills needed to effectively and creatively develop instruments that comprehensively assess students’ competency outcomes (Drisko, 2014).

This article focuses on one form of evaluation, providing an illustrative guide for social work educators who seek to develop comprehensive examinations to assess students’ acquisition of...
professional competencies, particularly at the master’s level. While others have argued the objective or relative merits of comprehensive assessments (e.g., Bogo, Rawlings, & Johnson, 2013; Patterson, 2006), our aim is to assist those who have chosen to use this mode of student-competency assessment by (a) reviewing best practices associated with assessment development, (b) integrating key concepts and principles into a unified framework, and (c) illustrating that framework by evaluating our own experience with assessment development. We conclude with suggestions for social work educators and ideas for future innovation and progress in this area. We begin by providing some relevant background information and context.

**Social work competencies and student assessment**

A shift toward competency-based social work education as outlined in the CSWE (2008) EPAS moved the focus of accreditation, and hence curriculum design, away from content and structure and toward student learning outcomes. Moreover, a “competency-based approach refers to identifying and assessing what students demonstrate in practice” (CSWE, 2015, p. 6). The CSWE (2015) identified nine social work competencies, each of which is listed in Table 1.

In addition to explaining social work competencies, the EPAS (CSWE, 2015) presents a number of important concepts associated with the assessment of student learning outcomes. First, assessment should consist of the systematic gathering of data about students’ acquisition of social work competencies at generalist and specialized levels of practice. Second, competence should be perceived as holistic, reflecting features of student performance, knowledge, values, affective reactions, critical thinking, and appropriate exercise of judgment. Third, assessment processes must be multidimensional and incorporate a variety of data collection tools that might vary by context. In addition to assessing student learning outcomes, assessment data can be used for a number of other important purposes, including the evaluation and modification of explicit and implicit curricula (CSWE, 2015).

Of the extant approaches to assessing students’ acquisition of social work competencies, the most commonly employed measures are field evaluations (Drisko, 2014), using field instructor assessments of students’ competence in direct practice and macro practice settings (e.g., Bogo et al., 2002; et al., 2012). Other forms of student assessment include classroom-based evaluations, such as capstone projects; course assignments, presentations, and examinations; and portfolios or research theses (Drisko, 2014). Although common in the fields of education and medicine, relatively less focus in social work has been placed on the development of comprehensive examinations designed to assess students’ acquisition of social work competencies at generalist and specialized levels. As one exception, the 2013 CSWE Annual Planning Meeting featured a workshop on designing objective structured clinical examinations to assess student social work competencies (Bogo et al., 2013). Another exception is a study in which a comprehensive, Web-based examination was administered to final-semester MSW students in response to a university’s graduation requirements (Patterson, 2006). Although this particular instrument was novel, the assessment questions were composed on the basis of focal curriculum content areas rather than on social work competencies per se. Building on these developments, and in the spirit of multidimensional

<table>
<thead>
<tr>
<th>Competency</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Competency 1</td>
<td>Demonstrate ethical and professional behavior</td>
</tr>
<tr>
<td>Competency 2</td>
<td>Engage diversity and difference in practice</td>
</tr>
<tr>
<td>Competency 3</td>
<td>Advance human rights and social, economic, and environmental justice</td>
</tr>
<tr>
<td>Competency 4</td>
<td>Engage in practice-informed research and research-informed practice</td>
</tr>
<tr>
<td>Competency 5</td>
<td>Engage in policy practice</td>
</tr>
<tr>
<td>Competency 6</td>
<td>Engage with individuals, families, groups, organizations, and communities</td>
</tr>
<tr>
<td>Competency 7</td>
<td>Assess individuals, families, groups, organizations, and communities</td>
</tr>
<tr>
<td>Competency 8</td>
<td>Intervene with individuals, families, groups, organizations, and communities</td>
</tr>
<tr>
<td>Competency 9</td>
<td>Evaluate practice with individuals, families, groups, organizations, and communities</td>
</tr>
</tbody>
</table>

*Source. CSWE (2015).*
assessments, ongoing efforts to explore the effective development of comprehensive assessment instruments to measure social work student competencies are warranted.

**Developing assessments: Guiding principles and best practices**

In addition to the CSWE’s (2015) EPAS, a number of highly relevant principles and best practices are available to guide efforts to develop comprehensive assessment instruments (see Table 2 for an integrative framework, which we discuss further in the next section). For one, the “assessment triangle” perspective indicates that the process of constructing a competency assessment should include (a) a definition of the constructs to be assessed (i.e., cognition), (b) the development of appropriate measurement items for eliciting competencies (i.e., observation), and (c) the pursuit of valid inferences based on assessment data (i.e., interpretation; Pellegrino, Chudowsky, & Glaser, 2001, p. 44; Zlatkin-Troitschanskaia et al., 2015). Said another way, competency assessments should include “a defining model of student cognition and learning in the domain, a set of beliefs about the kinds of observations that will provide evidence of students’ competencies, and an interpretation process for making sense of the evidence” (Pellegrino et al., 2001, p. 44).

McClelland (1973) also offered an influential perspective on assessing competence, arguing that assessments should examine competencies comprehensively. As articulated later by Drisko (2014), students’ acquisition of social work competencies should be indicated by a multitude of measurement indicators rather than by any single indicator of a particular competency. McClelland (1973) also emphasized that tests of competencies should be designed to highlight change in individual learning or development over time. Thus, competency assessments could be issued to students at the beginning, midpoint, or end of an academic program.

<table>
<thead>
<tr>
<th>Table 2. Stages of developing a comprehensive examination to measure student social work competencies.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Planning</strong></td>
</tr>
<tr>
<td>Generate a clear explication of the constructs to be assessed</td>
</tr>
<tr>
<td>Ensure the assessment and its targeted constructs align with institutional goals, accreditation standards, and professional competencies</td>
</tr>
<tr>
<td>Engage stakeholders for planning</td>
</tr>
<tr>
<td>Create a written plan</td>
</tr>
<tr>
<td>Establish an explicit and well-advertised time line for implementation</td>
</tr>
<tr>
<td>Assign leadership roles, delegate tasks, and create a system of accountability</td>
</tr>
<tr>
<td>Determine the best format and delivery system for the assessment</td>
</tr>
<tr>
<td>Determine how assessment results will be interpreted</td>
</tr>
<tr>
<td>Identify needed resources</td>
</tr>
<tr>
<td><strong>Creating and Implementing</strong></td>
</tr>
<tr>
<td>Fulfill leadership roles and task delegation</td>
</tr>
<tr>
<td>Ensure assessment creators acquire the necessary knowledge and skills</td>
</tr>
<tr>
<td>Ensure adequate resources are available</td>
</tr>
<tr>
<td>Construct the assessment instrument</td>
</tr>
<tr>
<td><strong>Evaluating</strong></td>
</tr>
<tr>
<td>Subject the assessment to pilot testing</td>
</tr>
<tr>
<td>Include methods for evaluating the process of assessment delivery</td>
</tr>
<tr>
<td>Identify appropriate methods for analysis and analyze data</td>
</tr>
<tr>
<td><strong>Revising</strong></td>
</tr>
<tr>
<td>Engage stakeholders and other experts for revising measurement items</td>
</tr>
<tr>
<td>Use pilot test data to guide revisions</td>
</tr>
<tr>
<td>Consider various causes of problematic measurement items</td>
</tr>
<tr>
<td>Finalize revisions and prepare assessment for future use</td>
</tr>
<tr>
<td><strong>Perpetuating</strong></td>
</tr>
<tr>
<td>Promote and advocate an institutional culture in which evidence-based decision making and student assessment is continually valued and pursued</td>
</tr>
<tr>
<td>Maintain plans for ongoing assessment and instrument improvement</td>
</tr>
<tr>
<td>Determine when and how many times students will be assessed during their time in the program</td>
</tr>
</tbody>
</table>
Banta et al. (2009) present additional principles for effectively planning and implementing outcome assessments. Good assessment planning involves engaging stakeholders, including faculty members, student affairs professionals, academic administrators, and possibly students. Assessments should also be planned with respect to relevant institutional or organization goals, which might include accreditation standards and professional values. Creating a written plan is another way to bolster effective assessment planning. Written plans could include details relating to who is responsible for various tasks, how tasks are to be completed, and when tasks should be finalized. The planning of an assessment should also address timing so that the introduction of an assessment can be properly anticipated by prospective respondents and those charged with implementing the assessment.

Promoting change in general, and emphasizing assessment in particular, can be met with institutional or organizational resistance, and strong leadership can help ensure that implementation is successful (Banta et al., 2009). In addition to the establishment of a clear leader, various staff and faculty should assume core roles in the assessment process. This can provide helpful task delegation and accountability. It is important to note that sufficient resources are requisite to the successful implementation of outcome assessments. Such resources might include institutional budgets, incentives for assessment participation among prospective respondents, and time and materials needed to help assessment creators or implementers develop needed skills. Indeed, all those involved in the assessment process should be properly educated with respect to best practices in relation to assessment design and writing the assessment items. In addition to a focus on student outcomes, assessment instruments can be implemented in a way that captures information about the process of assessment delivery (i.e., process evaluation). This might take the form of student feedback following the completion of the assessment or discussion among stakeholders regarding suggestions for improvement (Patterson, 2006). Banta et al. (2009) also note that efforts to assess student outcomes are best sustained when “planning and implementation take place in an atmosphere of trust and within a culture that encourages the use of evidence in decision making” (p. 8).

An understanding of measurement science and best practices are also essential for developing assessments. DeVellis (2012) provides steps associated with effective scale development. Although these steps pertain most closely to creating measures of a single and specific latent construct, some of them are useful in the context of competency assessment in social work education. In one step, assessment developers should first determine clearly what they intend to measure. This determination can be informed by theory and past research, or in the case of social work education and accreditation standards, professional competencies. Then a pool of items can be generated, the number and form of which depends on the situation or goals at hand. DeVellis (2012) also advocates for the involvement of experts who can help create or review items to be used as part of a measurement instrument. This helps ensure the content validity of an assessment instrument or that the full range of targeted content is covered by measurement items. Further, assessment items should be subjected to pilot testing, allowing subsequent revisions or adjustments to items as needed. Finally, data from initial tests of an assessment should be subjected to empirical evaluation. The form of evaluation will depend on the type of data collected (e.g., qualitative, binary-response items, or continuous-response items), but could include an estimation of item means, standard deviations, total scores, and the percentage of items answered correctly across respondents. Depending on the data collected, rigorous evaluations of item performance could be conducted using confirmatory factor analysis or item response theory models (DeVellis, 2012).

**An integrative framework and case example**

The principles and best practices summarized here informed our construction of an integrative framework for the effective development of comprehensive assessment instruments that measure
student social work competencies. We now focus our attention on the development of examinations (although we might use the terms assessment and examination interchangeably), which our framework presents in five stages: planning, creating and implementing, evaluating, revising, and perpetuating. Table 2 summarizes key tasks associated with each stage. We note that these stages are not exhaustive. We also note that the stages do not necessarily unfold in a strictly linear sequence. Indeed, tasks at one stage might overlap or coincide with tasks at the others. Ultimately, the integrative framework we present and illustrate in this article should be viewed as a general guide for the effective development of comprehensive examinations that measure social work student competencies.

We cast the summary of our case example on the backdrop of the integrative framework in an effort to evaluate our own process of examination development and to illustrate key principles and tasks in actual practice. Our case example took place at a long-standing school of social work at a research-intensive public university in the Southeast region of the United States. The school employs 30 tenure-track and 60 fixed-term faculty, and it enrolls about 125 students per year, delivering the curriculum through full- or part-time face-to-face classes.

The process of examination development focused on MSW students and their acquisition of social work competencies at generalist and specialization levels of practice. This particular MSW program houses two core concentrations: (a) direct practice and (b) community, management, and policy practice (i.e., macro practice). Typically, two thirds of enrolled students choose the direct practice concentration, with the remainder choosing the macro practice concentration. In the following, we focus on the process of assessment development rather than the specific outcomes or particulars of our assessment and its items.

Planning

To begin assessment planning, seven faculty were assigned to a steering committee, and three academic administrators served in ex officio capacities. The committee was chaired for the duration of the process by a faculty member with past accreditation experience. Over a 12-month period, in partnership with the associate academic dean and in consultation with all program faculty, the steering committee evaluated options for competency measurement and presented its recommendations to the full faculty for approval, resulting in the selection of field evaluations and comprehensive examinations as the school’s measures. The responsibility for field evaluation was assigned to the field education faculty, whereas the steering committee focused largely on the examination, creating a written process and specific operating assumptions for employing comprehensive examinations. Over the next 12 months, the chair guided the committee in the explanation of constructs to be assessed, communicated with students and stakeholders about the assessment, and identified the time line and needed resources. Numerous collaborative meetings took place in which assessment format, length, and delivery methods could be discussed. Moreover, the steering committee continually referred to accreditation standards, institutional goals, and social work competencies to guide planning efforts. Members of the committee were delegated tasks and given a timetable for task completion. The committee, under the leadership of the committee chair, also determined how the assessment results would be interpreted.

Ultimately, the decision was made to form two separate comprehensive examinations: one to establish a generalist-level understanding of social work competencies and another to establish a specialization-level understanding of advanced competencies. In terms of format and delivery, both assessments were conceptualized as multiple-item, multiple-choice assessments designed to cover each of the nine social work competencies, accessed using Qualtrics, a Web-based survey delivery system (Patterson, 2006). Our decision to use multiple-choice items was informed on several fronts. For one, we intended to mimic the formatting of the Association of Social Work Boards licensing exams. Thus, the assessment could be used to assess students’ acquisition of professional competencies and to familiarize students with the formatting of the board’s exams.
Second, measurement scholars have noted that multiple-choice items can be used to assess effectively a broad range of knowledge and comprehension and higher order thinking skills (Haladyna, 1997; Rodriguez, 2005).

Creating and implementing

Regularly scheduled meetings helped ensure that leadership roles were being fulfilled and tasks were being completed. Moreover, all faculty received training and practice in constructing multiple-choice test items and were tasked with item development. In the initial preparation of the examination item pool, all faculty were encouraged to generate items, regardless of whether they taught the classes that served as the basis of the examination. Administrators, faculty with research appointments, and those teaching elective courses were paired with professors teaching in the required MSW curriculum to brainstorm items. As an inducement to build a robust item pool, two $25 gift cards to the university bookstore were awarded in a drawing based on the number of usable items each faculty member generated. This incentive also acknowledged item construction efforts as an additional demand on faculty, above and beyond other tasks and meetings required to prepare for the self-study and reaffirmation. For the population of the specialization item pool, and for the continued development of the generalist item pool, faculty in particular curricular areas were responsible for item generation linked to the competencies associated with their courses.

Generalist items were constructed in a way that incorporated various stakeholders’ views and substantive expertise. Because there was strong alignment between the school’s mission and the explicit curriculum, faculty were also able to generate items that captured well the school’s mission and culture. Each item was assessed in terms of its explicit link to a generalist-level understanding of a core competency and refined by the steering committee and course instructors as needed prior to pilot testing. The initial pool of questions for the generalist assessment included 144 items, which were imported into Qualtrics. In an effort to balance adequate coverage of each competency with respondent burden, a random allocation of four items per competency was to be administered to respondents for pilot testing, resulting in a total of 36 foundation assessment items for each respondent. This approach would help ensure that all constructed items would be pilot tested.

The following year, faculty constructed specialization-level competencies and the associated examination, using an updated revision of Bloom’s taxonomy as a guide to target higher level cognitive processes (Krathwohl, 2002). The specialization examination used case-based, multiple-choice questions that were drafted by faculty and refined by an individual with specialized expertise in item construction. These case vignettes, and their associated questions, were used to capture nuances associated with social work practice that would help demonstrate students’ attainment of advanced social work competencies. A total of 60 specialization competency questions were constructed: five items each for Competencies 1 through 5 and for Competency 9 (30 items), five items each for Competencies 6 through 8 with a focus on direct practice (15 items), and five items each for Competencies 6 through 8 with a focus on macro practice (15 items). Thus, depending on their concentration, MSW students would be differentially assigned items for Competencies 6 through 8, with a total of 45 items administered to each student. We note that differences in item pool sizes between the generalist and specialization examinations should not be interpreted as an asymmetric focus on one examination versus the other. Rather, the difference reflects, in part, the relatively greater difficulty in composing specialization case vignettes and related items that capture higher level and concentration-specific thinking. Similar to the generalist examination, we limited the number of specialization items per competency in an effort to balance adequate coverage of each competency with respondent burden. Table 3 displays illustrative examples with two items from each examination.

The generalist and specialization exams both included an introductory statement, specifying the anonymity of responses and providing the explicit purpose of the assessment. Items included captured each respondent’s year of study and student status (e.g., full-time at main campus,
advanced standing, distance education). In terms of fonts, text sizes, and text spacing, the default settings in Qualtrics were used, which demonstrated adequate legibility. For the generalist examination, sets of items were allocated in the order of the social work professional competencies (i.e., one through nine, sequentially). For the specialization examination, items were organized by case vignettes. Thus, items reflecting various competencies could be allocated in a set of items linked to a case vignette. However, concentration-specific items were linked to unique sets of case vignettes, one set for direct practice and another for macrolevel practice.

**Evaluating**

As the assessments were being constructed, the steering committee determined the time and place the assessments would be administered to students for pilot testing. Initial efforts were focused on the creation and evaluation of the generalist assessment. All MSW students were asked to participate in the generalist examination pilot test at the end of the spring 2015 semester. Proctored sessions were scheduled in classrooms and the building’s computer lab for local and full-time students, and proctored sessions were similarly scheduled in other locations for distance education students.

After data were collected, members of the steering committee reconfirmed the desired form of outcome information that would be most useful; the primary result of interest was the percentage of correct responses, in total and for each competency. There was also interest in examining these percentages by program (i.e., full-time, advanced standing, local and satellite part-time cohorts) and year of study (i.e., first year, final year). Also estimated was the mean number of items answered correctly, in total, and per competency (along with standard deviations). The analysis also identified items for which 100% of responses were correct, or 60% or less of responses were correct. This was designed to help identify items that were potentially too easy or too difficult. Student feedback was also reviewed to identify issues with the delivery of the assessment or the quality of items.

### Table 3. Example items from the generalist and specialization examinations.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Competency</th>
<th>Example Item</th>
</tr>
</thead>
</table>
| **Generalist Examination** | Competency 1 | It is permissible to release client information without written consent when:
- a) A family member of a deceased client requests information
- b) The client poses a risk to herself/himself or someone else
- c) A social worker coordinates care with non-agency professionals
- d) A third party payer needs information to process a claim

As a social worker, many of your clients are of a different race or ethnicity from your own. It is important to:
- a) Acknowledge the difference with clients
- b) Be aware of the difference yourself but say nothing to the clients
- c) Ignore the difference since it doesn’t matter
- d) Discourage the client from talking about the difference if they bring it up

| **Specialization Examination** | Competency 1 | Associated Vignette: Roger is a homeless teenager. He has told Alex, his social worker at the shelter, that he is HIV positive. Roger and his girlfriend have histories of intravenous drug use and trading sex for cash. They frequently stay at the shelter and usually get one meal a day at the local food pantry.
- Which of the following steps would not be an appropriate action by Alex?
  - a) Explore Roger’s access to medical care
  - b) Encourage Roger to talk about his concerns
  - c) Discuss this case in supervision
  - d) Report Roger’s diagnosis to the public health department in order to get him help

Alex, the social worker, overhears another worker (case worker) arranging a pregnancy test for Roger’s girlfriend. What should Alex do first?
- a) Confirm with the case worker about the situation in a scheduled team meeting
- b) Ask Roger how he feels about fatherhood
- c) Place an anonymous call to the girlfriend to get an HIV test
- d) Call both clients’ parents

| **Competency 1** | | |
A similar process was later used to create, pilot test, and evaluate the specialization examination, using lessons learned during the creation and evaluation of the first pilot test of the generalist examination. To pilot test the specialization examination, a group of recently graduated MSW students were asked to complete the examination during the spring 2016 semester. The results allowed us to ensure that assessment items performed adequately. The respondents were also able to provide feedback related to the delivery of the assessment or the quality of items using a postassessment, open-response survey question.

**Revising**

The steering committee reviewed student feedback and examination results. Particular attention was paid to potentially problematic items, that is, items that might have been too easy or too difficult. After review, the committee and faculty members made adjustments to examination questions and response options. In some cases, items were omitted when it was determined that they performed poorly as a result of not being adequately congruent with course content or social work competencies. It is worth highlighting the nuances and challenges associated with this process of revision. Indeed, poorly performing items could stem from multiple causes, for example, the item was poorly written, the item had a poor or confusing set of response options, or the item was not adequately congruent with curriculum content or social work competencies. Poorly performing items could result from one or a mix of these issues. Thus, the committee was careful to attend to these possibilities in meetings and discussions. Following revisions, examinations were prepared for future formal administrations. For consistency and congruence with the school’s performance benchmarks, the random allocation of items per competency in the generalist examination was raised to five, resulting in a total of 45 items for each respondent on the generalist and the specialization examinations.

**Perpetuating**

Preparations for the formal administration of assessments coincided with the school’s year of record for reaccreditation. Foundation MSW students were administered the generalist-level examination, and advanced or final year MSW students were administered the specialization-level examination at the end of the spring 2017 semester. Students accessed examinations using Qualtrics on their personal laptops, tablets, or smart phones while in attendance at various sections of a specific required course. Thus, the instructors for this course served as proctors for the examinations. If needed, students could attend a proctored session in their program’s computer lab. In addition, discussions among stakeholders commenced to determine other contexts in which the administration of examinations might be desirable. For example, one decision included administering the generalist-level examination to future cohorts of incoming MSW students so that student social work competencies could be monitored over time and in relation to the completion of generalist-level course work. Because the assessments were designed to collect students’ feedback following the completion of the assessments, this form of process evaluation can continue to inform future assessment revisions and indicate assessment alignment or misalignment with the program curriculum and social work competencies.

The steering committee reviewed the results of the examinations and facilitated a full faculty retreat in spring 2017 to discuss implications for admissions, advising, the explicit curriculum, and other elements of the program. Although the steering committee disbanded at the conclusion of the reaffirmation process, the leadership for continuity of the evaluation process was assumed by the academic associate dean and related curriculum committees.
Conclusions and future directions

We present a number of conclusions associated with our integrative framework and process of assessment or examination development. For one, our case example was highly congruent with best practices, and key tasks to be completed at various stages of assessment development were illustrated well; however, we note that selecting the proper format of an assessment instrument can be arduous. Fortunately, previous work supports our selection of a Web-based, multiple-choice assessment format (e.g., Haladyna, 1997; Patterson, 2006; Rodriguez, 2005).

We also note that the process of developing an assessment is time consuming and often met with challenges. One clear challenge is coordinating schedules with already busy faculty to plan, create, and implement an assessment. Summer breaks and frequent changes in teaching assignments present considerable obstacles to this coordination. In addition, a great deal of variability is likely to emerge with respect to faculty knowledge and skill in creating valid and reliable assessment items. It is also difficult to properly train, engage, and incentivize faculty and other stakeholders in the process of writing items and to ensure congruence with course material without inadvertently enabling teaching to the test. Other challenges include ensuring that multidimensional competencies are adequately covered in an assessment, deciding and justifying how many items per competency to administer to students, handling missing data properly in the context of varying item allocation strategies and analysis, and effectively evaluating the psychometric properties of binary examination items (e.g., correct vs. incorrect response) in relation to underlying substantive constructs (e.g., a specific social work competency).

We offer several recommendations on how work in this area could best move forward to address these challenges. With respect to the planning, creating, and implementing stages of assessment development, it is vital to have adequate lead time to identify the linkages between the curriculum and the dimensions of each competency to be tested and to create, pilot-test, and revise the assessment. The steering committee must be well informed about the EPAS and the measurement plan so they can consider options and make informed recommendations to the full faculty. The steering committee chair serves as process manager shuttling between the committee and the faculty and item writers who teach the courses aligned with the competencies. We relied on faculty expertise in measurement, instrument construction, item writing, and statistical analysis to assist in carrying out these tasks and assist other faculty to cultivate expertise in these areas. One alternative under consideration is to contract with a person who has item-writing expertise to work with faculty content experts to ensure a robust pool of items for the examinations and a thorough reflection of the dimensions in each social work competency. Clerical assistance was essential to track changes in items at every step of the process, ensuring that the most current version of the examinations was available at all times and that any changes conformed to the measurement intent of a given item.

The initiative also required, and benefited from, extensive discussions about what performance on the examinations would mean and how students would be informed about the examinations and the results. The steering committee decided, and the faculty agreed, that examination performance would not be tied to graduation, although students would not be administratively cleared for graduation unless they completed the examination. Construing examination performance as a reflection on the curriculum and faculty, not on the students, was congruent with the aims of the measure and the uses of the results. The language used in the school avoided the word examination or exam, identifying the assessments as the Generalist Accreditation Assessment, GAA, and the Specialization Accreditation Assessment, or SAA. These decisions dissipated student anxiety and resistance and allowed the examinations to be administered later in the academic year and closer to graduation. Students who aspired to take professional licensure or certification exams found the examinations to be an opportunity to practice translating their education into a fixed-response format. Administering the examinations in the classroom rather than e-mailing the examination Web link and allowing students to complete it at their own pace helped to concentrate students’ attention on the examination, increase completion rates, and protect the item pool from...
Unauthorized distribution or misappropriation. The use of proctors and inclusion of the university’s honor code in the examination sign-in processes also aided in discouraging cheating.

In addition, we attended to examination duration and response patterns as indicators of students’ superficial engagement with the examinations. Moving forward, after bolstering confidence in examination performance, the steering committee will also consider providing direct feedback to students and their advisers on performance, or at least to those with poor performance, referring them to resources for improvement in relevant content areas. Although we found students to be fairly self-motivated, these options could further motivate students to put forth their best effort when completing the examinations.

In terms of challenges revealed during the evaluating, revising, and perpetuating stages of examination development, extant analytical methods might be helpful. For one, traditional confirmatory factor analysis incorporates continuous indicators of a latent or unobserved construct and assesses the extent to which the hypothesized construct accounts for the variation and covariation among observed indicators. In the context of binary-response items, such as those with either correct or incorrect response outcomes, one can assume the existence of a continuous response distribution that underlies each item response (Edwards, Wirth, Houts, & Xi, 2012). A respondent’s score along the continuous, latent construct spectrum is hypothesized to dictate which of the two response categories a respondent will produce (e.g., correct or incorrect). The point along the latent construct at which a respondent answers the item correctly is known as the item threshold. Confirmatory factor analysis with binary items incorporates a tetrachoric correlation matrix to estimate associations between binary items and model parameters (Edwards et al., 2012). This type of analysis could be used to evaluate the psychometric properties of items that are hypothesized to represent students’ understanding of a particular competency—the underlying latent construct.

Similar to confirmatory factor analysis with binary items, the item response theory framework provides a selection of models that are well suited to evaluate the process by which individuals respond to items (Edwards, 2009). Specifically, the two-parameter logistic model is appropriate for binary-response items. This model estimates item slopes, or discrimination parameters, and thresholds, or difficulty parameters. Slope parameters contain information about how related an item is to the latent construct being assessed. More specifically, slope parameters indicate how well a response to an item discriminates between respondents who receive one binary-response outcome versus the other. Higher slopes indicate a stronger ability of an item to discriminate. The threshold parameter indicates where along the latent-construct continuum being measured the respondent has a 50% chance of endorsing one binary-response outcome versus the other (Edwards, 2009). Higher threshold values indicate greater item difficulty or the need to possess higher levels of the latent construct to endorse the correct item response (Edwards, 2009). In all, item response theory models could also be used to effectively evaluate the psychometric properties of items used to measure student social work competencies.

Depending on the process used to deliver items to respondents, an understanding of missing data theory and best practices for handling missing data might be essential. Enders (2010) presents a comprehensive overview of principles associated with handling missing data, particularly regarding use of full information maximum likelihood and multiple imputation procedures. Both approaches are considered state of the art (Enders, 2010). Various software packages are well equipped to incorporate these approaches, including Stata Statistical Software (Version No. 14) and Mplus (https://www.statmodel.com/).

At this point, we also want to highlight some possible next steps for expanding our framework. By way of validating examinations, student scores from examinations could be correlated with other student outcome assessments, such as field-instructor ratings and grades or performance indicators from students’ required courses. These correlations would provide a sense of an examination’s criterion validity (DeVellis, 2012). In terms of predictive validity, students’ examination scores could be linked to postgraduation professional performance on the basis of alumni and employer surveys. Examination developers should also examine the extent to which examination items
demonstrate measurement invariance across students with respect to cultural background. Methods in confirmatory factor analysis and item response theory frameworks could be applied to meet these ends (Edwards, 2009; Millsap & Olivera-Aguilar, 2012)

Overall, our experience and integrative framework represent a starting point, not a finish line. The concepts and process we present should not be considered universally prescriptive but should encourage social work educators to think creatively about feasible methods for evaluating student competencies. Faculty members embedded in very small programs with a scarce labor supply or who deliver unique curricula structures are especially encouraged to apply our framework creatively. Moreover, the development of a curriculum-specific, Web-based, multiple-item, multiple-choice comprehensive assessment is one of many potential ways to effectively measure student social work competencies at generalist and specialization levels of practice. In this regard, we echo Drisko’s (2014) sentiment:

Analyzing what competencies are and how they are optimally assessed can improve social work’s efforts to ensure graduates are fully qualified beginning professionals. Defining competencies clearly and creating valid methods to assess them is vital to strong professional education. This is an important undertaking for the profession, as well as to serve and to protect the public. (p. 425)

In an era of student outcome assessment, our hope is that this framework and example will help guide other social work educators in the development of comprehensive examinations aimed to measure students’ acquisition of social work professional competencies.

Notes on contributors

Todd M. Jensen is research associate and Kimberly J. Strom-Gottfried is the Smith P. Theimann Distinguished Professor for Ethics and Professional Practice at University of North Carolina at Chapel Hill.

References


