FSEHD Advanced Programs Unit Assessment System

Design, Validity, and Consistency

September 2011

Contents

Advanced Programs Unit Assessment System

Learning Targets

In spring 2005, the faculty, led by the Associate Dean for Graduate Programs and the FSEHD Dean began in earnest to develop a common, coherent, comprehensive assessment system for advanced programs. Based on the agreed-upon Advanced Competencies, and drawing upon assessments already valued within programs (e.g. admission and midpoint interviews with students to assess dispositions and set goals and capstone performances such as comprehensive examinations, masters theses, and portfolios) the unit developed assessments that were aligned to the Advanced Competencies and to develop common rubrics for program and unit evidence collection. Since 2008, the Advanced Competencies been revised, and a major summative unit assessment has been implemented. The remainder of the Advanced Programs Unit Assessment System is currently under review.

In spring 2005, faculty members who coordinate advanced programs worked together, with guidance from the dean's office, to develop a common Advanced Programs Assessment System for the unit. The first step was the development of advanced competencies that are linked directly to the unit's Conceptual Framework. Advanced competencies based on the Four Themes of the Conceptual Framework were articulated to guide instruction and assessment. These included Knowledge (General Knowledge, Domain-Specific Knowledge, Technology Knowledge); Practice (Communication and Expression; Reflective Problem-Solving; Professional Practice; Technology Use); Diversity (Systems View of Human Development; Family Centeredness and Engagement; Individuals Differences and Cultural Diversity); Professionalism (Professional Ethics; Collaboration; Leadership; Professional development).

These original Advanced Competencies were closely aligned with the unit's existing Professional Dispositions (see Table 2).

Table 2: Alignment of Dispositions to Advanced Competencies

Advanced Competencies	***FSEHD Unit Dispositions***
Knowledge	Self Reflection Life-Long Learning
Practice	Collaboration Professional Work Characteristics
Diversity	Advocacy for Children and Youth Respect for Diversity
Professionalism	Collaboration Professional Work Characteristics Life-Long Learning

Based on data analyses and suggestions from faculty, the Advanced Programs Coordinators committee revised the unit's advanced competencies in 2008. The goal of this endeavor was to make the Advanced Competencies more meaningful and relevant to the diverse programs at the advanced levels. Ultimately the process resulted in narrowing several existing advanced competency categories, adding clearer language consisting of demonstrable verbs, and re-configuring the topical headings from four to two. Knowledge and Practice are now the larger headings with Diversity and Professionalism infused throughout them, the idea being that what any candidate knows (Knowledge) and can do (Practice) must be in the context of Diversity and Professionalism. The revised Advanced Competencies include Professional Awareness, Information Literacy, Contextual Perspective, and Domain-Specific Knowledge in the category of Knowledge (infused with Diversity of Professionalism) and Evidence-based Decision Making, Technology Use, Diversity of Practice, and Professional Identity Development in the category of Practice (infused with Diversity of Professionalism). The revised Advanced Competencies are illustrated in Table 1.

Table 1: Advanced Competencies (Revised 2008)

Knowledge influenced by diversity and professionalism *FSEHD advanced candidates demonstrate the requisite knowledge of content and practice to prepare them to be experts of the diverse fields of their disciplines.*	**Practice informed by diversity and professionalism** *FSEHD advanced candidates incorporate their domain-specific knowledge into performance with attention to diversity and the standards of their profession.*
Knowledge 1.) Domain-Specific Knowledge: candidate demonstrates conceptual mastery of subject matter, literature, theory, and methods in one's chosen field of professional practice.	Practice 1.) Evidence-based Decision Making: candidate defines a problem clearly; collects/analyzes data; uses data to inform decision-making; addresses target population dynamics; and incorporates considerations of other professionals and/or stakeholders while determining a plan of action that: a) contributes to school improvement and/or renewal; and/or b) promotes the well-being of children, family systems, school systems, or communities.
Knowledge 2.) Information Literacy: candidate recognizes when information is needed and has the ability to locate, interpret, and evaluate relevant information.	Practice 2.) Technology Use: candidate selects and uses technology effectively in: a) presentation of information, b) collaborative work environments, c) information collection analysis and management, and d) research based activities
Knowledge 3.) Contextual Perspective: candidate demonstrates a comprehensive understanding of diversity as it relates to field specific content.	Practice 3.) Diversity of Practice: candidate uses knowledge of diversity about self and others to design effective practice.
Knowledge 4.) Professional Awareness: candidate exhibits an understanding of the standards of one's chosen profession, (e.g., confidentiality, ethics)	Practice 4.) Professional Identity Development: candidate examines own emerging, developing or acquired professional knowledge, skills, communication, and dispositions that will result in competent practice, and creates plan to further one's own professional growth.

While Diversity and Professionalism influence all Advanced Competencies, each one has been determined to align most closely with the themes of the Conceptual Framework as follows:

Table 2. Alignment of Conceptual Framework and Advanced Competencies

Conceptual Framework	*FSEHD Advanced Competencies*
Knowledge	Domain Specific Knowledge; Information Literacy
Pedagogy	Evidence-Based Decision Making; Technology Use
Diversity	Contextual Perspective; Diversity of Practice
Professionalism	Professional Awareness; Professional Identity Development

Guiding Principles

> "…a collection of assessments does not entail a system any more than a pile of bricks constitutes a house. Therefore, the fundamental question for school leaders is: In what sense does their plan constitute a *system* of assessments, rather than a *collection* of assessments?" (Coladarci, 2002, p. 773)

Assessment systems are clearly made up of individual assessments. Yet, a collection of individual assessments is not considered an assessment system unless they are guided by a "coherent plan for assessment" (Coladarci, 2002, p. 773). The following six features distinguish an assessment *system* from a collection of assessments and were used to guide the development of the Advanced Programs Assessment System and the plan for its implementation:

1. The assessments collectively are relevant to announced learning targets.
2. The assessments are conducted at multiple time points (e.g., admission, mid-point (formative), exit, and post-graduation)
3. Each assessment has an announced purpose
4. The system is made up of assessments that are initiated at multiple levels (e.g., self vs. external evaluation, classroom vs. program vs. unit levels)
5. Candidates are allowed multiple opportunities to demonstrate knowledge, understanding, and skill development
6. The assessments draw on multiple formats—"traditional" and "alternative" alike (Coladarci, 2002, pp. 73-74; Maine Comprehensive Assessment System Technical Advisory Committee, 2000, pp. 3-4)

These six characteristics also advance the validity of inferences made on the basis of advanced unit assessment results. The current status of the six characteristics in terms of design and implementation within the FSEHD Advanced Programs System are described below:

The assessments are relevant to announced learning targets.

The Advanced Programs Assessment System was specifically designed to provide evidence of student achievement of the learning targets specified earlier in this document: the Advanced Competencies (linked to the unit's Conceptual Framework), the Unit Dispositions, and the Culturally Competent Teaching Areas. The targets of each component of the assessment system are public, and the rubrics/criteria for judging student performance on each learning target are explicit. The learning targets for each assessment are printed on the assessment and the accompanying rubric. Finally, the unit is working toward improving the measurability of the learning targets relevant to each unit assessment. To this end, FSEHD has developed, field tested, and is implementing new common rubrics for unit data collection and has plans to review and potentially revise the entire unit assessment system for advanced programs.

Each assessment has an announced purpose

The Advanced Programs Assessment System has been explicitly designed to make clear the purpose each assessment has within the system. Each assessment within the system serves one of the following purposes:

- Admission: Evaluation of candidate qualifications to enroll in an FSEHD advanced programs
- Formative: Evaluation of candidates as they proceed through the programs, identifying weaknesses in candidates and programs so student remediation or program improvements can be offered in a timely fashion. This is implemented prior to student teaching at the initial teacher preparation level. It is typically implemented mid-program or prior to an extended field experience or internship at the advanced level.
- Summative: Evaluation of candidates at the end of an advanced program to ensure that applicants and candidates are qualified to graduate and to identify strengths and weaknesses of the programs and the unit
- Post: Evaluation of candidates for program and unit evaluation

In addition, there are four primary audiences for each assessment. Assessment data can be utilized to address questions and concerns relevant to students, faculty, program coordinators, and unit staff. Examples of questions and concerns pertaining to various audiences over the four transition points include (but are not limited to):

- Candidates: Am I improving over time? Am I succeeding at the level that I should be? What help do I need?
- Faculty: Does this candidate meet the admissions or exit criteria for our program? Which candidates need help? What grades should candidates receive? Are my instructional strategies working?
- Program: Is our program effective? How can it be improved? Which candidates are making adequate progress? Are our candidates ready for the workplace or the next step in learning?
- Unit: Who is applying to our programs? Are programs producing the intended results? How should we strategize to achieve success? Which programs need/deserve more resources? (Stiggins, 2001, pp. 11-12)

The assessments are conducted at multiple time points

The Advanced Programs Assessment System includes four checkpoints where knowledge, skills and dispositions are assessed: admission, formative (e.g., mid-point), exit, and post-graduation. This allows program and unit staff to monitor candidate progress toward mastery of relevant learning targets. Each program has identified a course of action if assessments indicate that candidates are not yet ready to proceed to the next stages of their programs is being adopted by each program. Possible actions include remediation, re-doing assessments, and denial of advancement. Documentation of these instances is being kept by all programs.

The four transition points are currently in place for all programs. At the Advanced levels, the admission transition point occurs at the stage when a candidate applies to a program. The formative assessment transition point for initial teacher preparation programs is just prior to student teaching. The faculty in each advanced program have identified a formative checkpoint that is appropriate to the nature and structure of its particular program. Some timeframes programs have chosen to identify the formative point include: the mid-point in terms of number of credits in the program, after 12 credits of graduate study (application for candidacy), prior to an internship or clinical experience, etc. The summative transition point occurs at the during the student teaching experience for initial teacher preparation candidates. Summative unit assessments are implemented at the end of advanced programs but are sometimes split over the two final semesters.

The post-graduation checkpoint is fully implemented, with the administration of graduate follow up surveys to FSEHD alumni of Advanced programs and surveys of graduates' employers. This process was first initiated in 2005/2006 and repeated in 2010/2011. Data have been analyzed, and recommendations for program improvement are available. Additionally, while all advanced program graduates are currently taking required professional licensure or certification exams and obtaining valid certificates for their profession, the assessment system is not yet equipped to collect, analyze, or report on this data.

The system is made up of assessments that are initiated at multiple levels

According to the Standards for Educational Accountability Systems established by the Center for Research on Evaluation, Standards, and Student Testing, assessment systems "include data elements that allow for interpretations of student, institution, and administrative performance" (Baker et al., 2002, p. 2). Including assessment data from multiple levels (e.g., classroom, program, unit, etc.) facilitates the process of identifying areas of improvement in each area (American Education Research Association, American Psychological Association, & National Council on Measurement in Education, 1999).

Consequently, the assessments in the Advanced Programs Assessment System are initiated at the individual course (I), program (P), unit (U), and state or national (SN) levels. As displayed in the assessment system blueprints (Table 1), GPA and minimum course grades are derived from course specific assessments (I); however, many courses also include assessments that are common across a particular program (P) or the unit (U). The work sample required at the advanced formative assessment point are program specific (P), yet assessed with a unit wide rubric (U). The PIP used at the summative point in Advanced programs is initiated at the program level (P), yet assessed with a unit wide rubric (U). On the other hand, candidate self evaluations and faculty evaluation of candidates are initiated at the unit level (U). Furthermore, state or national level professional licensure certification exam results (SN) are utilized to provide additional information regarding the achievement of Advanced program graduates. Graduate follow up and employer surveys are initiated at the unit level (U) but are disaggregated at the program level (P). The use of multiple measures allows for the assessment of students, programs, and the unit through multiple lenses and allows for the triangulation of evidence used to make inferences about student achievement and program effectiveness. This, in turn, increases the validity of such inferences.

Candidates are allowed multiple opportunities to demonstrate knowledge, understanding, and skill development

The design of the advanced program assessment systems afford candidates multiple opportunities to demonstrate their growth in the learning targets identified by their programs and the unit. Additionally, the use of common learning targets, criteria, and rubrics as candidates progress through their programs clarifies expectations and enables faculty and candidates to observe candidate growth as they participate in multiple opportunities to demonstrate their knowledge, understanding, and skill development over time. The use of multiple assessments with multiple formats, as opposed to a single, "one-shot" assessment, increases the validity of the inferences subsequently made regarding the knowledge, skills, and dispositions of advanced programs candidates.

The assessments draw on multiple formats—"traditional" and "alternative" alike

There are many methods for assessing learning; yet, no single assessment format is adequate for all purposes. (American Educational Research Association, 2000) Consequently, the advanced program assessment system allows candidates to demonstrate their knowledge, skills, and dispositions using a variety of methodologies. The various assessment methodologies used in the Advanced Programs Assessment System are classified as follows:

- *Selected Response and Short Answers*: Assessments that ask candidates to choose from pre-selected responses, such as multiple choice, true/false, or matching questions. Short answer questions are also included here. These assessments are a good match for evaluating content knowledge and to a lesser extent for the application of knowledge to solve problems.
- *Constructed Response*: Assessments that require substantial responses that candidates construct for themselves on paper. Included here are essays, graphic representations, case studies, and other ways for candidates to demonstrate their knowledge and skills on paper. This method of assessment is often a good match for evaluating content knowledge and the application of knowledge to solve problems.
- *Performance Tasks*: Assessments that require candidates to provide evidence of their knowledge or skills by demonstrating them "in the moment" or by creating artifacts that are similar to those created by professionals in their area of interest. Included here are projects, presentations, and exhibitions. This method is a good match for evaluating candidates' skills as practitioners in their field.
- *Observation and Personal Communication*: Assessments that classroom faculty carry out as part of their daily teaching and assessment repertoire as they observe and communicate with candidates, including formative assessments such as check lists, anecdotal records, conferencing, journal entries, and guided conversations. This method also includes candidate self-evaluation, as candidates reflect on their experience and learning and evaluate their own strengths and weaknesses. This method is a particularly good match for evaluating the dispositions of candidates. (Smith & Miller, p. 17)

As shown in the Advanced Programs Assessment System blueprint (Table 1), all four assessment formats are utilized throughout the four assessment transition checkpoints. This attempt to "balance" assessment in terms of assessment methods yields multiple forms of diverse and redundant types of evidence that can used to check the validity and reliability of the judgments and decisions. (Wiggins, 1998)

Design

The Advanced Programs Unit Assessment System is in a time of flux, or transition. Because the is currently under review and is likely to be revised substantially in 2011-2012, the system is presently operating under two overlapping sets of learning targets: the original Advanced Competencies (adopted in 2005)and the Revised Advanced Competencies (developed and adopted in 2008). At this time, only the newly designed PIP is aligned with the revised Advanced Competencies and revised Professional Dispositions. The other formative and summative assessments are aligned with the "old" or original Advanced Competencies and dispositions. As they were not designed with an eye toward the revised learning targets and they are likely to be eliminated, the unit has not undertaken efforts to align the existing Summative Assessments to the newly revised Advanced Competencies or the revised Professional Dispositions.

	ADMISSIONS	ORIGINAL ADVANCED COMPETENCIES	FORMATIVE	SUMMATIVE	POST
Reflective Practice	**Grade Point Average** **Standardized Test Score** **Professional Goals Essay** **Performance-based Evaluation** **Candidate Reference Form**	**KNOWLEDGE** • METACOGNITIVE KNOWLEDGE • DOMAIN-SPECIFIC KNOWLEDGE • TECHNOLOGY KNOWLEDGE	Minimum grade of B in courses/ assessments tied to standards	Exit GPA and comprehensive assessment	Professional Licensure/ Certification Exam Results (as applicable)
		PRACTICE • COMMUNICATION & EXPRESSION • REFLECTIVE PROBLEM-SOLVING • PROFESSIONAL PRACTICE • TECHNOLOGY USE	Individual work sample in key course/practicum	_***Professional Impact Project (PIP)—<u>aligned with revised Advanced Competencies</u>***_	Valid certificate (as applicable); Surveys of Graduates (biennial) & Employers (triennial)
		DIVERSITY • SYSTEMS VIEW OF HUMAN DEVELOPMENT • INDIVIDUAL DIFFERENCES & CULTURAL DIVERSITY • FAMILY-CENTEREDNESS & ENGAGEMENT	Candidate's self-reflection of progress Faculty's reflection of candidate progress	Candidate's self-evaluation of outcome Faculty's evaluation of candidate achievement	Surveys of Graduates (biennial) & Employers (triennial)
		PROFESSIONALISM • PROFESSIONAL ETHICS • COLLABORATION • LEADERSHIP • PROFESSIONAL DEVELOPMENT			

The blueprints for the Advanced Programs Assessment Systems is displayed in Table 1 below.

Table 1: Advanced Programs Assessment System Blueprint

KEY	**Methods**	**Level**	**Status**

	SR=selected response/short answer; CR=constructed response; PA=performance assessment; OC=observation/ personal communication	I=individual course;
		P=program;
		U=unit;
		SN=state or national
			E=existing;
			P=planned

ADVANCED PROGRAMS

Transition Point	**Assessment**	**Method**	**Level**	**Status**
Admission	GPA	SR, CR, PA, OC	I	E
	Standardized test score	SR	SN	E
	Professional goals essay	CR	U	E
	Performance based evaluation	PA	U	E
	Candidate reference form	OC	U	E
Formative	GPA (minimum B average); assessments tied to standards	SR, CR, PA, OC	I	E
	Work sample	CR, PA	P, U	E
	Self-evaluation*	OC	U	E
	Faculty evaluation	OC	U	E
Summative	GPA	SR, CR, PA, OC	I	E
	Comprehensive assessment	CR	P	E
	Professional Impact Project	PA	P, U	E
	Self-evaluation*	OC	U	E
	Faculty evaluation	OC	U	E
Post Graduation	Professional Licensure/ Certification Exam (as applicable)	SR, CR	SN	E
	Valid certificate (as applicable)	varies	SN	E
	Graduate follow up survey*	OC	U	E
	Employer survey	OC	U	E

*Note: The only self-evaluations in the Advanced Programs Assessment System are the self-evaluations at Formative and Summative points and the Graduate follow up survey after graduation.

As shown in the table above, the assessment systems include four checkpoints at which knowledge, skills and dispositions are assessed: admission, formative (Preparing to Teach for initial program and Formative for advanced programs), summative (Exit for initial programs and Summative for advanced programs), and post-graduation. In addition, the systems are aligned with the Conceptual Framework, RIPTS, Culturally Competent Teaching Areas, and Advanced Competencies, as appropriate. Assessments within the systems occur at multiple levels, and include multiple measures of student achievement. Finally, all but three assessments at the post-graduation transition point have been implemented as of this point.

Transition Points

Admissions

Candidates applying for admission to an advanced program at FSEHD are required to meet the following criteria for acceptance:

- Teaching certificate (for all school related programs except school psychology).
- Official standardized test scores (Graduate Record Examination or the Miller Analogies Test) required for all FSEHD graduate degree plans except CGS in Physical Education.
- Three candidate reference forms completed by former instructors, employees, or other professionals who can assess the candidate's potential to complete graduate study and make a positive influence in the field. Candidates are evaluated on the following dimensions: capacity for insight, clarity of goals, intellectual curiosity, motivation and initiative, rapport with children and youth, rapport with adults, emotional stability, adaptability to change, reliability and dependability, ability to organize ideas
or tasks, oral and written communication skills, and overall potential.
- Professional goals essay, essay, including the candidate's reflection on experiences, skills, and lifelong learning; level of preparation, knowledge base, and professional activities; professional goals and their relation to serving all individuals and families; and reasons for choosing the FSEHD advanced program.
- Performance-based evaluation (i.e., a recent teaching or work-performance evaluation).
- Other program specific requirements.

Formative

Each advanced program has identified a formative checkpoint that is appropriate to the nature and structure of the particular program. Some timeframes programs have chosen to identify the formative point include: the mid-point in terms of number of credits in the program, after 12 credits of graduate study (application for candidacy), prior to an internship or clinical experience, etc. To progress past the Formative Transition Point, advanced candidates must demonstrate successful performance on each of the following requirements:

- A minimum Grade Point Average of B or better
- A Performance-Based Work Sample in a key course or practicum. The appropriate work sample for this transition point is determined by the program and is intended to provide evidence of the candidate's growing competency in the Advanced Competency, Practice. In particular, the work sample must display the candidate's skills in Communication and Expression, Reflective Problem-Solving, Professional Practice, and Technology Use. The work sample is evaluated using a four-point, unit wide rubric.
- Self-evaluation: The candidate assesses the extent to which s/he is developing a series of attributes/behaviors related to FSEHD's Advanced Competencies and the Unit Dispositions since his/her admission into the advanced preparation program. The evaluation instrument consists of a 12-item, four-point Likert-type rating scale.

- Faculty evaluation: A faculty member assesses the extent to which the candidate is developing the same attributes/behaviors as described above. The evaluation instrument contains the same items in the same format as the self-evaluation.

Note: The Advanced Programs Assessment System is presently being reviewed and is likely to be revised substantially in 2011-2012. For this reason, the existing formative assessments are aligned with the "old" Advanced Competencies and dispositions. As they were not designed with an eye toward the revised Advanced Competencies or the revised Professional Disposition and they are likely to be eliminated, the unit has not undertaken efforts to align the existing Formative Assessments to the newly revised Advanced Competencies or the revised Professional Dispositions.

Summative

To exit their programs, advanced and program candidates must demonstrate successful performance on each of the following requirements:

- A minimum Grade Point Average of B or better
- Successful performance on the program's Comprehensive Assessment
- Professional Impact Project. FSEHD advanced candidates are required to complete FSEHD Professional Intervention Project (including descriptive rubrics) for Advanced Programs (PIP) to be completed by at the end of their programs. The purpose of this assessment is for advanced program candidates to create a relevant Professional Intervention Project for Advanced Programs that includes all Practice aspects of the revised Advanced Competencies: Evidence-Based Decision Making; Technology Use; Diversity; Professional Identity Development. Through this Professional Intervention Project process, it is expected that advanced candidates will provide credible evidence of their ability to facilitate impact on constituents and reflect upon their practice. Several faculty members volunteered to field test the PIP in Spring 2009. Based on their feedback, the PIP was revised and subsequently implemented on the unit level in Fall 2010. The current status of PIP implementation is displayed in Table 3.

Table 3: Status of PIP Implementation in Advanced Programs

Program	Course implemented	When	Material submitted to Assessment
Advanced Studies in Teaching and Learning	SED 555: Literacies across the disciplines. Teacher research project for the capstone experience.	Spring	This spring, 2011, will be the first time that data will be submitted on the new PIP rubric.
Early Childhood	ELED 510 + ELED 662	Part of the final project, a pilot research study done in ECED 662. The course will next run in Fall 2011.	No, not available yet.
Educational Leadership	LEAD 505 LEAD 505/511	Spring 10 Spring 11	No – beginning Spr 11
Elementary Education	ELED 664 or FNED 547	Fall 2011	No – beginning fall 2011
Health Education	**HED 505** Program	Spring 2011	

	Development	Complete needs assessment, implementation plan, & evaluation plan	
	HED 562 Seminar in Health Education	Fall 2011 Implement program, conduct evaluation, analyze data, write report , submit report	No – beginning fall 11
Mental Health counseling	CEP 683/684	Fall 10/spr 11	No – beginning spr 11
Reading	ELED 663		
School Counseling	CEP 541/542	Fall 10/spr 11	No – beginning Spr 11
School Psychology	CEP 629	Spring 11	No – beginning spr 11
SPED Early Childhood	SPED 648	Fall 10	Yes
SPED Exceptional Learning Needs			
SPED Initial cert			
SPED Severe/Profound			
SPED Urban Multicultural			
Teaching English as Second Language	FNED 547 (ELED 510)	Fall 10 (LB) & spring 2011 (JJ)	C & W

Other advanced summative assessments parallel those administered at the formative stage and include:

- Self-evaluation: The candidate assesses the extent to which s/he has developed a series of attributes/behaviors related to FSEHD's Advanced Competencies and the Unit Dispositions since his/her admission into the advanced preparation program. The evaluation instrument consists of a 12-item, four-point Likert-type rating scale.
- Faculty evaluation: A faculty member assesses the extent to which the candidate has developed the same attributes/behaviors as described above. The evaluation instrument contains the same items in the same format as the self-evaluation.

SPA Assessments

Each nationally recognized FSEHD advanced program includes six to eight SPA assessments among their program assessments. Many SPA requirements are measured using the same instruments as - and simultaneously with – the unit's outcomes assessments. Where SPA assessments differ from unit assessments, programs design and administer SPA assessments themselves, collect and analyze program-level assessment data, and present these data during SPA reviews.

Post-Graduation

The post-graduation checkpoint consists of employer and graduate follow up surveys. Employer surveys are administered electronically to school and district administrators in the state of RI, as well as directors of community agencies that are likely to employ FSEHD advanced program graduates. The employer survey asks respondents to compare the caliber of recent FSEHD advanced graduates to that of advanced graduates from other advanced preparation programs. The survey also asks respondents to assess graduates' mastery of the Advanced Competencies and their overall advanced program preparation and included opportunities for open-ended feedback regarding their perceptions of the strengths and weaknesses of advanced preparation at FSEHD. The most recent Employer Survey was administered in November 2010.

Graduate follow up surveys are administered electronically to all FSEHD advanced graduates who have successfully exited an advanced program within the past five years. Graduates are contacted via email. Additionally, links to the survey are posted on relevant RIC sites on Facebook. Follow up surveys of advanced program graduates require graduates to rate their mastery of the Advanced Competencies and their overall FSEHD advanced preparation and include opportunities for graduates to provide open-ended feedback regarding the strengths and weaknesses of their programs and their overall experiences at FSEHD. The most recent graduate follow up survey was administered in February 2011.

Procedures for Ensuring That Key Assessments of Advanced Candidate Performance Are Fair, Accurate, Consistent, and Free of Bias

FSEHD believes that assessment accuracy, consistency, and freedom (and other concepts) from bias are key components of validity and reliability. The procedures FSEHD uses to ensure that key assessments of candidate performance are fair, accurate, consistent, and free of bias are, as well as findings from sample validity and reliability studies, are presented in the sections below.

Validity

Fairness, accuracy, and freedom from bias are among the key components of validity. FSEHD utilizes a comprehensive approach to define and examine the validity and utility of its assessment system and the inferences that the system yields. Validity is a multifaceted concept and the most important technical consideration for assessments and assessment system design and use. Furthermore, the notion of usefulness, or utility, of assessment results is inherent in any examination of validity. Critical components of validity are being monitored in the Advanced Programs Assessment System. These include the six principles underlying the Advanced Programs Unit Assessment System, as well as:

- Content-related validity: Do assessment items/components adequately and representatively sample the content area(s) to be measured?
- Construct validity: Do assessments and the assessment system measure the content they purport to measure?
- Prediction: How well do assessment results predict how well candidates will do in future situations?
- Fairness: Are all candidates afforded a fair opportunity to demonstrate their skills, knowledge, and dispositions?
- Utility: How useful are the data generated from unit assessments?
- Consequences: Are assessment uses and interpretations contributing to increased student achievement and not producing unintended negative consequences? (Linn, 1994)

Content-related validity. With regard to content-related validity, assessments are aligned with the original and revised FSEHD Advanced Competencies and dispositions identified by the larger professional community as important. Additionally, the new PIP assessment was designed based on best practice in higher education assessment as identified in the literature and on the professional knowledge, experience, and consensus of FSEHD faculty, many of whom are developers and definers of best practice in their professional areas.

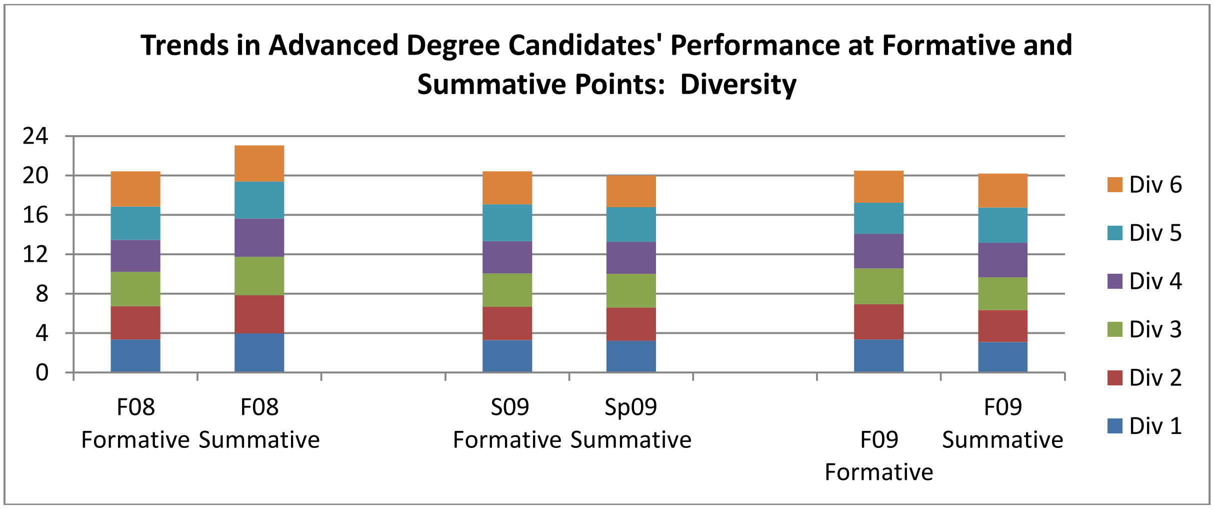
The Advanced Programs Unit Assessment System is in a time of flux, or transition. Because the is currently under review and is likely to be revised substantially in 2011-2012, the system is presently operating under two overlapping sets of learning targets: the original Advanced Competencies (adopted in 2005)and the Revised Advanced Competencies (developed and adopted in 2008). At this time, only the newly designed PIP is aligned with the revised Advanced Competencies and revised Professional Dispositions. The other formative and summative assessments are aligned with the “old” or original Advanced Competencies and dispositions. As they were not designed with an eye toward the revised learning targets and they are likely to be eliminated, the unit has not undertaken efforts to align the existing Summative Assessments to the newly revised Advanced Competencies or the revised Professional Dispositions.

Construct-related validity. An assessment has construct validity if it accurately measures a theoretical, non-observable construct or trait. The construct validity of an assessment is worked out over a period of time on the basis of an accumulation of evidence. FSEHD is investigating the validity of its unit assessment and accumulating validity evidence in a number of ways. High internal consistency is one type of evidence used to establish construct-related validity. That is, if an assessment or scale has construct validity, scores on the individual items/indicators should correlate highly with the total test score. This is evidence that the test is measuring a single construct. Mean internal consistency coefficients of scales included in the summative faculty and self-evaluations of advanced candidates and the summative Capstone assessment in the past three years is adequate and range from .70 to .85 for the constructs of Diversity, Professionalism, and Practice, as evaluated by faculty and advanced candidates themselves. These findings provide some evidence for the construct validity of the constructs being assessed the faculty evaluation, self-evaluation, and Capstone assessment.

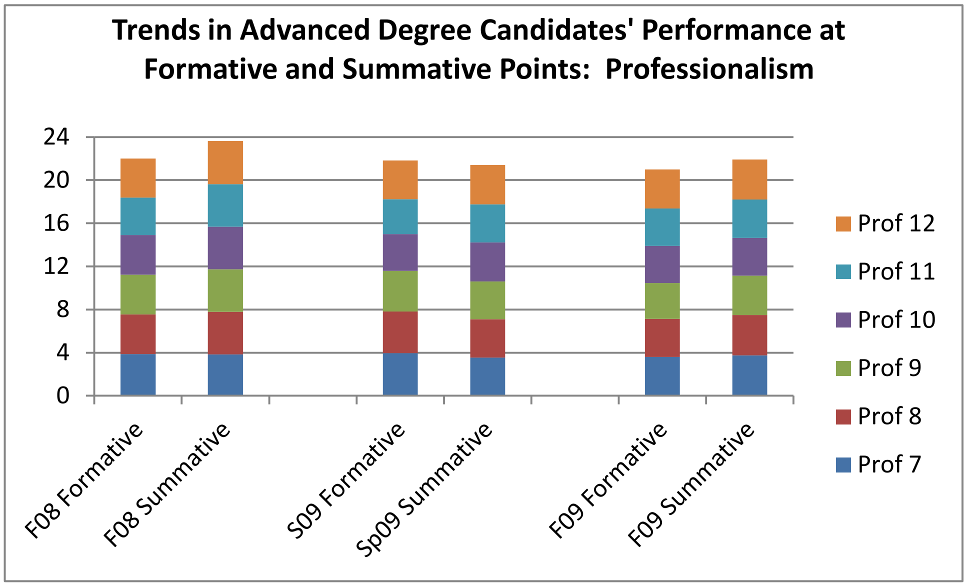
Second, FSEHD has conducted factor analyses to examine whether the theoretical framework of FSEHD unit advanced assessments match the factor representation yielded by confirmatory factor analysis. A confirmatory factor analysis of Spring 2009 candidate self-evaluation data (principal components extraction, varimax rotation, n=115) revealed a four factor solution accounting for 61% of the variance in the data. These findings were interesting, considering that the instrument was designed to just two constructs: Diversity and Professionalism. Similarly, confirmatory factor analysis of faculty evaluation data revealed a three factor solution accounting for 69% of the variance in the data. In contrast, a factor analysis of Capstone assessment data reveal that the instrument is measuring a single construct (i.e., Practice), just as hypothesized by the unit. Together, the results of these analyses do not provide evidence to support the construct validity of the faculty and self-evaluation instruments. Conversely, data support the construct validity of the Capstone assessment.

Another method of collecting construct-related validity evidence is to examine developmental changes. Assessments measuring certain constructs can be shown to have construct validity if the scores on the tests show predictable developmental changes over time. FSEHD has investigated the construct validity advanced unit assessments by examining whether the developmental changes expected to occur from the Formative to Summative stages are actually evident. Specifically, two sets of data were analyzed for this purpose: 1) faculty ratings of advanced candidates at the formative and summative stages were examined over three semesters to investigate the construct validity of the 12 item, Likert rating scale faculty evaluation instrument (targeting the Advanced Competencies of Diversity, Practice and Professionalism) and 2) faculty evaluations of candidates from the identical rubrics used in the Formative Work Sample and Summative Capstone Assessment. Analyses of the candidate evaluation instrument allowed for investigation of expected developmental growth in the Diversity and Professionalism, while analyses of the Work Sample and Capstone assessments examined developmental change in the Advanced Competency of Practice.

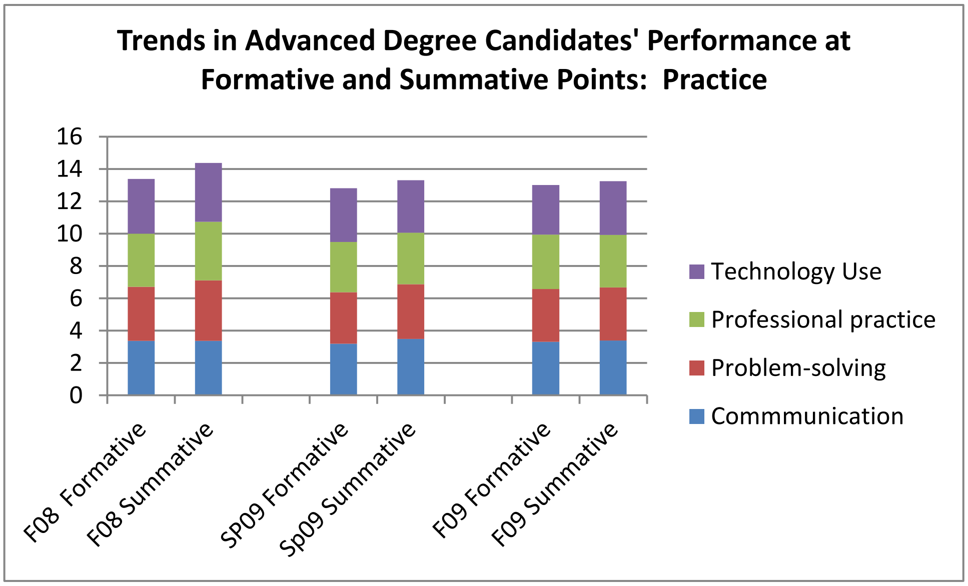
With the exception the Fall 2008 semester, there is little change in faculty ratings between the formative and summative assessment points in the diversity competency. This lack of variability suggests limited construct validity on the part of the faculty evaluation instrument.



Data reflect maintenance or small increases between the formative and summative assessment points in regards to the Professionalism competency. Again, these data raise questions about the construct validity of the faculty evaluation instrument.



When examining Practice data, slightly larger increases between formative and summative assessment points are observed. Increases exist in all practice areas: communication, problem-solving, professional practice, and technology use.



In conclusion, these analyses do not provide strong evidence of the construct validity of the faculty evaluation instrument and suggest that FSEHD faculty would be well served by investigating assessment strategies that will yield greater validity evidence for advanced candidate growth in Diversity and Professionalism. On the other hand, data from studies of Practice data yield greater evidence of the construct validity of unit's measures of the Advanced Competency of Practice.

Prediction. FSEHD conducts ongoing checks of the predictive validity of assessments in the unit assessment system. While it is not feasible to investigate the predictive validity of every assessment each year, the unit conducts "spot checks" of predictive validity periodically. For example, the Director of Assessment conducted a study of the utility of standardized test scores (MAT, GRE) as admission criteria in the prediction of subsequent program performance was conducted. Results revealed that the MAT is highly and significantly correlated with GPA among advanced program non-completers. Among program completers, the MAT, GRE Verbal, and GRE Analytical tests are significantly correlated with GPA. These findings were used to support the recommendation that the standardized testing requirement be retained at admission to an advanced program.

Fairness. The following components are included in the fairness criterion and contribute to the extent to which inferences and actions on the basis of assessment scores are appropriate and accurate.

- Freedom from bias: The language and form of assessments must be free of cultural and gender bias.
- Transparency of expectations: Assessment instructions and rubrics must clearly state what is expected for successful performance.
- Opportunity to learn: All candidates must have had learning experiences that prepare them to succeed on an assessment.
- Accommodations: Candidates with documented learning differences must be afforded accommodations in instruction and assessment.
- Multiple opportunities: Candidates must have the opportunity to demonstrate their learning in multiple ways and at different times. (Smith & Miller, 2003)

FSEHD has addressed the fairness criterion as follows:

As unit assessments have been designed and revised, the wording and design of assessment tasks were reviewed by the respective program faculty with a focus on whether the selected tasks were fair and accessible to all candidates. The design, format, wording, and presentation of unit assessments have been reviewed multiple times by internal and external constituents, and changes have been made in an effort to minimize any unintentional bias. For example, the new summative advanced program assessment, the Professional Impact Project, was renamed as such after faculty indicated that the previously suggested title, Professional Intervention Project, suggested a "deficit model" of education. The title and select other terms in the document were subsequently revised to reflect less biased language.

FSEHD also examines unit advanced assessment results for bias according to gender and ethnicity. This is done semester by semester at the admissions, formative, and summative stages. For example, the data presented below demonstrate very similar GPAs for males and females at the formative stage. In contrast, the GPA for Hispanic advanced students (3.76) is slightly lower than that of White students (3.87). The third set of data below show that the GPA for white males is slightly higher than that of all other gender/ethnic groups. Nevertheless, the GPAs for all sub-groups of students are quite high, with a minimum of 3.76 on a four-point scale. These and other findings findings do not suggest that unit advanced assessments are biased according to gender and ethnicity.

Cumulative GPA * Gender

Female	Mean	Median	Std. Deviation	Minimum	Maximum	N
Cumulative GPA	3.8600	3.9450	.18129	3.52	4.00	10

Male	Mean	Median	Std. Deviation	Minimum	Maximum	N
Cumulative GPA	3.8800	3.8800	.12000	3.76	4.00	3

Total	Mean	Median	Std. Deviation	Minimum	Maximum	N
Cumulative GPA	3.8646	3.9300	.16470	3.52	4.00	13

Cumulative GPA * Ethnicity

Hispanic	Mean	Median	Std. Deviation	Minimum	Maximum	N
Cumulative GPA	3.7600	3.7600	.	3.76	3.76	1

White	Mean	Median	Std. Deviation	Minimum	Maximum	N
Cumulative GPA	3.8733	3.9450	.16886	3.52	4.00	12

Total	Mean	Median	Std. Deviation	Minimum	Maximum	N
Cumulative GPA	3.8646	3.9300	.16470	3.52	4.00	13

NOTES:

Two ethnic groups represented in the data.

Gender * Ethnicity * Cumulative GPA

Male Hispanic	Mean	Median	Std. Deviation	Minimum	Maximum	N
Cumulative GPA	3.7600	3.7600	.	3.76	3.76	1

Total Hispanic	Mean	Median	Std. Deviation	Minimum	Maximum	N
Cumulative GPA	3.7600	3.7600	.	3.76	3.76	1

Female White	Mean	Median	Std. Deviation	Minimum	Maximum	N
Cumulative GPA	3.8600	3.9450	.18129	3.52	4.00	10

Male White	Mean	Median	Std. Deviation	Minimum	Maximum	N
Cumulative GPA	3.9400	3.9400	.08485	3.88	4.00	2

Total White	Mean	Median	Std. Deviation	Minimum	Maximum	N
Cumulative GPA	3.8733	3.9450	.16886	3.52	4.00	12

NOTES:

Equal white male and white female medians. Female White to Male White ratio 5:1.

FSEHD has also made numerous efforts to make assessment expectations transparent and to keep faculty, cooperating teachers, and candidates informed. First, the Director of Assessment and the assessment committee have strived to keep faculty and cooperating teachers up to date and informed on all aspects of the revised assessment system. Information was shared with them during faculty retreats in February 2007, August 2008, August 2009, February 2010, and March 2010. Information has also been shared with faculty through the Dean's Leadership Committee and electronic correspondence with faculty. Advanced assessments are also a frequent topic of discussion at monthly meetings of advanced program coordinators. Finally, the PIP rubrics and prompts are highly descriptive and provide detailed guidance to the candidate and evaluators about expectations and process. They also contain multiple, observable indicators that make expectations explicit and yield granular information about multiple dimensions of a candidate performance.

Unit assessment tasks were reviewed by the respective program faculty in terms of opportunity to learn and succeed at the content/skills inherent in the tasks during the assessment design and revision process. Furthermore, while the PIP has been piloted, the unit has acknowledged faculty members' willingness to take a risk with a new assessment AND conscious that the instrument was indeed being piloted. Hence, FSEHD is granting programs (and candidates) some leniency in these areas during the early pilot semesters. The unit did not set a cut off score for the PIP; rather, it provided faculty with general guidelines and then allowed faculty members to assign a PIP as passing or failing based on their professional judgment and the status of the program in relation to emphasizing certain TCWS concepts and skills. Additionally, faculty in FSEHD programs have begun to engage in curriculum mapping and other processes to examine what is taught at different time points to ensure that candidates have opportunities to learn and succeed at the content and skills inherent in unit assessments.

Rhode Island College is committed to making reasonable efforts to assist individuals with documented disabilities. Candidates seeking reasonable classroom or assessment accommodations under the ADA of 1990 and/or Section 504 of the Rehabilitation Act of 1973 are required to register with Disability Services in the Student Life Office. To receive accommodations for any class or unit assessment advanced candidates must obtain a Request for Reasonable Accommodations form and submit it to their professor at the beginning of the semester. This information is shared with all advanced candidates at the beginning of each course. In addition, this information is provided in each course syllabus.

Multiple opportunities to demonstrate learning and growth are built into the very design of the FSEHD unit advanced assessment system. The system includes many opportunities for candidates to demonstrate their learning—in multiple ways and at different times. Furthermore, candidates are afforded opportunities to retake or redo all or part of their unit assessments. The use of multiple assessments with multiple formats, as opposed to a single, "one-shot" assessment, increases the validity of the inferences subsequently made regarding the knowledge, skills, and dispositions of FSEHD candidates.

Consequences. Linn (1994) states, "it is not enough to provide evidence that the assessments are measuring intended constructs. Evidence is also needed that the uses and interpretations are contributing to enhanced student achievement and, at the same time, not producing unintended negative consequences." (p. 8) Positive, intended consequences of the FSEHD unit assessments include improved learning on the part of candidates/graduates, as well as program and unit improvement based on the use of assessment data. Negative, unintended consequences might include a narrowing of the curriculum (to focus on preparation for assessments) or increased student drop out due to unanticipated burdens of the Assessment System. Positive, unintended consequences of the system may occur, as well, and these should be identified.

Graduate follow up surveys of graduates are conducted at the "post" transition point and include opportunities for graduates to provide open-ended feedback regarding the strengths and weaknesses of their programs and overall experiences at FSEHD. This qualitative data has been analyzed for clues as to consequences of the assessment system. At this time, data from program graduates do not reveal any negative unintended consequences of the unit assessment system at the initial or advanced levels. Feedback from faculty regarding the functioning of the assessment system and the consequences thereof is always welcome at FSEHD and is solicited on an ongoing basis. Feedback regarding the positive and negative intended and unintended consequences of the Advanced Program Assessment System are gathered and reflected upon, and none reveal unintended negative consequences of unit assessment.

Multiple Measures. FSEHD unit assessments draw on multiple formats—"traditional" and "alternative" alike. There are many methods for assessing learning; yet, no single assessment format is adequate for all purposes. (American Educational Research Association, 2000) Consequently, the FSEHD advanced assessment system allows candidates to demonstrate their knowledge, skills, and dispositions using a variety of methodologies. The various assessment methodologies include: Selected Response and Short Answers; Constructed Response; Performance Tasks; and Observation and Personal Communication.

As shown in the, all four assessment formats are utilized throughout the four advanced assessment transition checkpoints at FSEHD. This attempt to "balance" assessment in terms of assessment methods yields multiple forms of diverse and redundant types of evidence that can used to check the validity and reliability of judgments and decisions. (Wiggins, 1998)

Reliability

As discussed above, it is essential that the School utilize assessment instruments and procedures that permit valid inferences regarding the competencies and dispositions of their advanced program candidates. Further, reliability, or the consistency of scores across raters, over time, or across different tasks or items that measure the same thing, is a necessary condition for validity. FSEHD recognizes detailed rubrics facilitate the reliability of scoring the selected tasks and facilitate faculty training in the scoring system. While current advanced rubrics were developed in alignment with relevant learning targets, the unit is in the process of examining and revising advanced unit rubrics. For example, the rubrics for the new PIP assessment at exit is highly detailed and descriptive, with clear expectations at each performance level. Inter-rater reliability will thus be enhanced finished the revision of existing assessment instruments.

Research has demonstrated that the reliability coefficients for teacher-made assessments generally range from .60 to .85 (Linn & Gronlund, 2000), while standardized tests of achievement and aptitude tend to fall between the .80s and low .90s (Salvia & Ysseldyke, 1998). Assessment experts agree that the required level of reliability for assessment increases as the stakes attached to the assessments increase (i.e., when assessment-based decisions are important, permanent, or have lasting consequences) (Linn & Gronlund, 2000). Salvia and Ysseldyke (1998) specify a minimum reliability of .90 for assessments that are used for tracking and placement. FSEHD is working toward a goal of reliability coefficients of at least .85, given 1) the seriousness of the decisions made about students based on assessments in the Advanced

Programs Assessment System and 2) the newness of many aspects of the assessment systems. With time, the school will aim toward even higher levels of reliability.

FSEHD also routinely conducts studies to establish consistency, or reliability, of assessment procedures. Beginning in spring 2007, assessment data have been analyzed, and coefficients of internal consistency and/or inter-rater reliability (depending on the assessment) have been computed to determine how well this criterion was achieved. Additionally, the Director of Assessment conducts Many-Facet Rasch Measurement analyses to address the following research questions for performance assessments and rating scales in the system:

- Do the scorers differ in the levels of severity they exercise, or do both groups of raters function interchangeably?
- Do faculty and advanced program candidates rate in the same manner?
- Are there any inconsistent raters whose patterns of ratings show little systematic relationship to the scores that other raters give?
- Do some advanced program candidates exhibit unusual profiles of ratings across scale dimensions, receiving unexpectedly high (or low) ratings on certain dimensions, given the ratings the candidate received on other dimensions?
- Are there any raters who cannot effectively differentiate between scale dimensions, giving each teacher candidate very similar ratings across a number of conceptually distinct dimensions?
- Are the categories on the rubrics and rating scales appropriately ordered? Are the rubrics and rating scales functioning properly? Are all the scale categories clearly distinguishable?

The computer program, FACETS (Linacre, 1988), is used to analyze the data and furnish answers to the research questions identified above.

The findings from these studies are used to refine/improve the Advanced Programs Assessment System, target faculty professional development needs, and serve as evidence of reliability of scoring and the validity of inferences made based on performance and rating scale assessments within the system. Investigations of this nature are conducted on an ongoing basis. Examples of these investigations and their results are presented in the following sections.

Inter-Rater Reliability Findings

On a day-to-day basis, inter-rater reliability at the advanced program level at FSEHD is enhanced by the collaborative nature in which faculty conduct their work and consider student progress. At the admissions, formative, and summative transition points in a candidate’s progress through the program, several reviewers consider materials. Programs have admission committees - generally 2 to 3 program faculty members - who independently review and score admission materials on a program admission rubric. These scores are transposed to the unit admission rubric as the final admission decision.

At the formative stage, assessment materials include a formative work sample, a candidate self evaluation, and faculty evaluation. These items are collected by the program faculty and submitted to the unit. Most programs have a formative review process conducted by program faculty that considers all of the candidate work to date (grades from classes, performance on portfolios, etc), and the formative stage assessment materials. All of these materials representing various perspectives on the candidate's performance are used to recommend continuation into next stages in the program (from Practicum to Internship, for example). If a candidate's work is evaluated as unsatisfactory at this stage, either a remediation plan is developed or the student is not recommended for continuation.

Similarly, at the summative stage of a candidate's program, a broad perspective is taken to determine candidate completion of program requirements. Information used to make final program completion decisions may include: a final portfolio or capstone assignment, field supervisor evaluation of performance, faculty supervisor's assessment of field performance, the Professional Impact Project, a candidate self evaluation.

An area in which FSEHD has computed the inter-rater reliability of scoring in advanced unit assessments is in the self and faculty ratings that take place at candidates' formative and summative stages. Each semester, correlations are run between Self Evaluation Ratings and Faculty Evaluation Ratings of Candidate Skills in Diversity and Professionalism to obtain measures of inter-rater reliability. Specifically, candidates' self-evaluations and faculty evaluations of candidates' developing skills in Diversity and Professionalism are compared to explore whether faculty and candidates were arriving at similar conclusions about the candidates' current status in these two areas. The following analyses are conducted:

- Correlations between self and faculty ratings are run to explore the degree of correspondence between the two sets of ratings. It is hypothesized that a shared understanding of the nature of the Diversity and Professionalism Advanced Competencies, as well as deep knowledge of the student on the part of the faculty, would be associated with high, statistically significant correlations between self and faculty ratings on the same indicators.
- Paired sample t-tests are conducted to determine whether mean self and faculty ratings on each Diversity and Professionalism indicator are statistically different from each other and not due to a chance finding ($p<0.05$). Cases where self and faculty ratings are significantly different potentially point to issues to address in advisement, instruction, and instrument refined
- Responses to similarly themed items are averaged to form new variables representing Mean Self and Mean Faculty Diversity ratings, as well as Mean Self and Mean Faculty Professionalism ratings. Correlations and paired sample t-tests are run to explore the relationships and differences between faculty and self ratings on these two variables

Inter-rater reliability findings in this area are generally disappointing, with inter-rater reliability coefficients (i.e., correlations) ranging from .00 to .73. Mean differences between faculty and self-ratings are also frequently statistically significant. These findings suggest that the rubrics on which faculty and candidates are rating candidates may not be sufficiently clear and that the two groups do not necessarily share a consistent view of what is expected of advanced candidates at the formative and summative stages. The faculty and self-evaluation instruments used in the FSEHD's advanced assessment system definitely warrant further inspection and refinement.

FSEHD's Director of Assessment also conducts inter-rater reliability analyses as requested by specific advanced programs. For example, a recent Many-Facet Rasch Measurement analysis of admissions data for the Educational Leadership program revealed that:

- All faculty evaluators scored with similar levels of severity
- There was no unwanted variation in rater severity that affected examinee scores
- All raters were internally consistent with regard to their application of the rating scales across all candidates and admission criteria
- The admissions criteria are well differentiated in terms of difficulty
- Candidates were differentiated in terms of their performance on admissions criteria
- Scores that candidates received on the admissions criteria were quite fair and not adversely affected by undue rater leniency or severity
- Admissions criteria target the competence levels of lead applicants well
- For all LEAD admissions criteria except the Professional Goals Essay, the scale categories are appropriately ordered and functioning properly.
- All rating categories are utilized with the Test Scores, Recent GPA, Experience, and Professional Goals Essay criteria. Raters are only able to discriminate two levels of difficulty for Performance Based Evaluation and Professional Goals Essay criteria, despite being required to use a four-point scale.
- Each rating category is contributing to meaningful measure of the particular admissions criteria.

Similarly, the Director of Assessment conducted a Many Facet Rasch Analysis of comprehensive exam results for the Counseling and Educational Psychology Programs. Findings revealed ways in which the programs could improve the assessment rubric and ensure greater consistency among scorers.

Internal Consistency Reliability Findings

FSEHD routinely studies the internal consistency of its advanced unit assessments. Recent studies have found the internal consistency of advanced assessments to be adequate. For example, mean internal consistency coefficients of scales included in the summative faculty and self-evaluations of advanced candidates and the summative Capstone assessment in the past three years is adequate and range from .70 to .85 for the constructs of Diversity, Professionalism, and Practice, as evaluated by faculty and advanced candidates themselves. Internal consistency of candidates' self-evaluations are consistently lower (.70-.75) than those of FSEHD faculty (.83-.89), suggesting a need to clarify expectations with candidates.

The studies, practices, and related improvements to the unit assessment system noted in the above section all lead to more consistent data collection across programs. The thoughtful, planned, and evidence based revision of the assessment system is leading to greater consistency in expectations and assessment practice across the unit. Emerging data indicate that the internal consistency of measurement with new assessments is quite high. Training of faculty and cooperating teachers is aimed at ensuring a common understanding of the design, purpose, and execution of unit assessments, as well as inter-rater reliability. Inter-rater reliability analyses are revealing areas in which further orientation, training, and calibration are needed. All in all, the unit is moving in a very positive direction with regard to developing greater consistency in data collection.

Standard Setting

A final technical criteria for a high quality assessment system is standard setting. In other words, programs and the unit must identify the amount and quality of evidence necessary to demonstrate proficiency on assessments. These are performance standards (*Measured Measures*, 2000). The standard setting process cannot begin until criteria for levels of student performance (i.e., rubrics) are well-articulated (Smith & Miller, 2003). This is why reliability must be established first.

FSEHD plans to train selected faculty in two approaches to standard setting by the end of 2011: the Angoff method and the Examination of Student Work method. The Angoff method is an assessment-based method for faculty to work collaboratively to determine a passing grade or acceptable performance on an assessment in a course. It can be used with traditional assessment types (such as selected response) that are frequently used on advanced coursework. The Examination of Student Work method of standard setting involves the review of student work (yielded through performance assessment) and results in the establishment of data-based cut off scores and anchor papers/benchmark performances. Skill in using these standard-setting methods and implementation of these procedures within programs will yield more consistent scoring of student work samples at the formative and exit transition points, as well as within courses in programs, resulting in higher reliability in candidate's final course grades. A subsequent training will be offered to faculty in 2012. Additionally, faculty who are trained in standard-setting methods will be encouraged to share their knowledge with their peers in their departments. Standard setting procedures will be applied to all unit performance assessments.

An additional, useful benefit of the standard-setting process is that it often exposes flaws in scoring rubrics or the design of assessments. As such, it is part of an iterative process of ongoing revision and improvement.

Next Steps Needed

Based on the results of validity and reliability studies and on feedback from faculty, the following steps are recommended to refine and improve the validity and consistency of advanced unit assessments:

- Continue to review and potentially revise the advanced programs assessment system. With a Director of Graduate Programs in place for the first time in two years, this task is now feasible. In particular, the unit should re-examine, revise, and/or eliminate the faculty and self-evaluations, which have demonstrated poor validity and reliability.
- Gather data on the utility of revised advanced unit assessments from faculty and candidates. As the advanced unit assessment system is revised, it will be crucial to study whether revised or new assessments are of utility to them as professionals and as trainers of future teachers.
- Engage in a formal standard setting process as soon as feasible.
- Continue to investigate the validity and consistency of the advanced unit assessment system on an ongoing basis.

References

American Education Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: Authors.

Baker, E., L., Linn, R. L., Herman, J. L., & Koretz, D. (2002). *Standards for educational accountability (Policy Brief 5).* Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.

Center for the Study of Evaluation & National Center for Research on Evaluation, Standards, and Student Testing. (1999). *CRESST assessment glossa*ry. Los Angeles, CA: CRESST/UCLA. Available: http://cresst96.cse.ucla.edu/CRESST/pages/glossary.htm

Haessig, C.J. & LaPotin, A.S. (2007). *Lessons Learned in the Assessment School of Hard Knocks*. Irving, CA: Electronic Educational Environment, UCIrvine. Available: http://eee.uci.edu/news/articles/0507assessment.php

Linacre, J.M. (1988). *FACETS*. Chicago: Mesa.

Linn, R. L., & Gronlund, N. E. (2000). *Measurement and evaluation in teaching* (8th ed.). New York: Macmillan.

Linn, R. L. (1994). Performance assessment: Policy promises and technical measurement standards. *Educational Researcher, 23* (9), 4-14.

McLeod, S. (2005). *Data-driven teachers*. Minneapolis: School Technology Leadership Initiative, University of Minnesota. Available at: www.scottmcleod.net/storage/2005_CASTLE_Data_Driven_Teachers.pdf

Measured measures: Technical considerations for developing a local assessment system. (2005). Augusta, ME: Maine Department of Education.

Smith, D. & Miller, L. (2003). *Comprehensive local assessment systems (CLASs) primer: A guide to assessment system design and use*. Gorham, ME: Southern Maine Partnership, University of Southern Maine.

Stiggins, R.J. (2001*). Leadership for Excellence in Assessment: A Powerful New School District Planning Guide*. Portland, OR: Assessment Training Institute.

Salvia, J., & Ysseldyke, J. E. (1998). *Assessment* (7th ed.). Boston: Houghton Mifflin.

Webb, N. L. (2005). *Issues related to judging the alignment of curriculum standards and assessments*. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Canada.

Wiggins, G. (1998). *Educative assessment*. San Francisco, CA: Jossey-Bass.