WILEY

Author Query Form

Journal   CC

Article   cc20359

Dear Author,

During the copyediting of your manuscript the following queries arose.

Please refer to the query reference callout numbers in the page proofs and respond to each by marking the necessary comments using the PDF annotation tools.

Please remember illegible or unclear comments and corrections may delay publication.

Many thanks for your assistance.

Query No.	Description	Remarks
Q1	Please confirm that forenames/given names (blue) and surnames/family names (vermilion) have been identified correctly.	
Q2	Please provide the reference corresponding to the citation "Cumming and Miller (2017)."	
Q3	Table 11.1 is given as a list format. Please advise if it could be set as a display list.	
Q4	The reference Zwick (2012) is listed in the references list but is not cited in the text. Please either cite the reference or remove it from the references list.	

11 *This chapter describes the role of faculty in conducting quality assessment and includes suggestions for preparing for successful meetings with reviewers on accreditation teams.*

Academic Assessment: Best Practices for Successful Outcomes with Accreditation Evaluation Teams

Tammie Cumming, M. David Miller

Accreditation is an important process for ensuring program integrity, as well as awarding Title IV financial aid, which includes student loans and federal grant awards. Assessment is a critical component of accreditation. Most accrediting commissions, including professional accrediting organizations, such as the ABET (formerly known as the Accreditation Board for Engineering and Technology), and the six regional accrediting organizations recognized by the Council on Higher Education Accreditation (CHEA), have requirements for a well-documented and resourced assessment process (Cumming & Zhao, 2015).

Accreditation standards across all CHEA-recognized regional accrediting bodies have in recent years become more rigorous with respect to student learning outcomes assessment, while extending the same standards to Administrative, Educational, and Student Support (AES) Units.

While this chapter focuses on academic learning, the reader should be mindful of the accreditation standards that have evolved to encompass all units.

Within this chapter, we present the foundational requirements for a quality assessment system that promotes faculty buy-in and presents examples from two different institution types. The University of Florida, a top-ranked public research institution provides examples of its advanced assessment system. The City University of New York–New York City College of Technology (City Tech), a highly diverse, comprehensive commuter college (formerly a community college) presents examples with a basic system that includes faculty in establishing the foundational building blocks of a quality assessment system. The University of Florida is accredited by

SACSCOC and received the highest level of commendation for its most recent accreditation event in 2016. City Tech is accredited by MSCHE and was also formally commended for its recent self-study accreditation action in 2018. Thus, we present two assessment systems that were considered best practices by these two regional accreditors.

Evolution of Assessment in Higher Education

Assessment has been an integral part of education for more than two centuries. The National Education Association reports that educators began formulating assessments for student achievement in 1838 (U.S. Congress Office of Technology Assessment, 1992). An early large-scale use of assessment was when Joseph Rice (1914) administered tests to thousands of students to examine the efficiency of the use of instructional time.

While assessment has had a long history in the public schools for K-12, it was slower to develop a central role in higher education. Ewell and Cumming (2017) suggest that the "assessment movement" in higher education could be dated to the First National Conference on Assessment in Higher Education in 1985 and the U.S. Department of Education's report, *A Nation at Risk* (1983). However, testing has become an important part of accountability throughout the education system from early childhood through higher education. Testing has been an integral tool in educational reform (Linn, 2003; Miller, Linn, & Gronlund, 2013), whether through national initiatives in K-12 (e.g., No Child Left Behind Act of 2001) or the increased testing requirements among the regional K-12 accrediting bodies in higher education.

With assessment maintaining a central role in education in all levels at this point, it is important to ensure the quality of assessments being used for accountability purposes. While assessment has long been an important part of teaching and learning in K-12, post-secondary education has focused on the unique expertise of the faculty teaching within a discipline and most likely has not included formal training in assessment. Thus, quality program assessment has been slower to evolve in post-secondary education. However, assessment practices and methodologies have become more sophisticated as the field has made some significant advancements.

These advancements were, in part, a result of faculty demanding a logical and scientific assessment system based on measurement principles since the investment in such activities can be substantial. Simply being assigned an assessment task within their university was understandably met with resistance. Faculty began asking questions regarding the quality of the assessments, sampling methodologies, and adequacy of the analyses that were conducted—and used—to make decisions that sometimes included a significant investment by programs with a potentially substantial impact. In order to address these issues, experts within the assessment and institutional research and effectiveness offices consulted the well-established

educational measurement literature to make advancements that would increase faculty buy-in and result in valid and generalizable results.

In the field of educational measurement and testing, the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME) have developed the *Standards for Educational and Psychological Testing* (*The Standards*) in 1985, which provides a framework for ensuring the quality of assessments in their development and use. The most recent version of *The Standards* (AERA, APA, & NCME, 2014) provides a framework for a solid assessment foundation that is applicable to assessments across a broad range of uses and contexts. As a result, many universities and colleges use *The Standards* as a guideline for ensuring the quality of their assessments, and ultimately increasing the confidence of faculty and administrators in utilizing this data.

The three foundational chapters of *The Standards* provide guidance for ensuring the quality of assessments and emphasize the importance of reviewing and establishing these properties at the onset of any assessment effort. First, *validity*, the evaluation of the adequacy and appropriateness of the interpretations and uses of test results, is considered a minimal requirement for any assessment, ranging from locally developed assessments to large standardized testing assessments. Second, *reliability*, the consistency of a measurement or how much random error there can be in the measurement, is discussed as a necessary but not sufficient condition to establish validity. This can include assessing consistency across the multiple items or tasks of an assessment, consistency in scoring an assessment, or consistency over time or forms. Third, *fairness*, ensuring that the test has the same properties across all types of people, can also be considered as a part of establishing validity. We would argue that effective assessment is an essential component of program (including general education) improvement and the three foundations should be considered as part of the assessment process from the initial planning of assessment to the final uses and interpretations. Examples are provided below in the three foundational areas of *The Standards*.

The Role of Faculty in Establishing Validity

Validation is the process of documenting the appropriate uses and interpretations of assessments. Validation could be documented with several types of evidence, ranging from the review of content alignment with respect to the instructional objectives/student learning outcomes, how examinees respond to the assessment, the internal structure of the assessment, the relationship to other measures, and the consequences of the assessment use. While validation can assume many methods and forms which vary in their level of sophistication, the most frequent method to document validity of student achievement is to document the alignment of

the content of the assessment with what it should be measuring; in higher education, it is establishing that an assessment measures content consistent with what is being taught and the specified student learning outcomes. This type of validity does not involve statistical or complex methods, but requires that "experts" review the content of the assessments and that the review is documented. These "experts" in a given content area are the faculty.

Establishing Validity for Program Assessments at the University of Florida

The University of Florida has formalized a content review process to ensure the validity of assessments through an examination of the alignment of the assessment with program goals and Student Learning Outcomes (SLOs). The Office of Institutional Assessment, whose mission is to establish, maintain, and refine institutional effectiveness and assessment processes is at the heart of the process by guiding the faculty to utilize best practices since the validation process really begins with program faculty. The assessment results not only provide the faculty with important information to evaluate their programs, it is also included in the SACSCOC accreditation report. Each academic program is required to complete an *Academic Assessment Plan*, which includes:

- A mission statement
- Program goals
- Student learning outcomes (SLOs)
- Curriculum maps
- Documented assessment cycle
- Methods, procedures, and measurement tools
- A sample rubric or locally developed exam used to measure an SLO
- Assessment oversight

In creating the *Academic Assessment Plan*, faculty must establish the content alignment between the assessment, the SLOs, and the curriculum. Thus, they are following the procedures for validation of assessment content.

Once the *Academic Assessment Plan* is completed, it undergoes a review by the Academic Assessment Committee (AAC). The AAC is a university-wide committee that provides oversight on academic assessment. This committee was primarily established to review the Academic Assessment Plans and the validity of the assessments being used by each program. The process of validation is expanded to include review by faculty not associated with the program to ensure that there is no bias in the process. The process of content validation includes a review of the assessment and its alignment with the SLOs and the Program Goals.

Consequently, without reference to validation or other assessment terminology, faculty for each program can participate in the validation process as a usual part of their curriculum and assessment review. The content review process as described includes content review by "experts" who are (1) the faculty in the content domain, and (2) faculty on a review committee specializing in examining assessments. In addition, the AAC will review the response processes as a second form of validation. Some program areas will voluntarily provide additional information about validation, including statistical analyses of the internal structure of the assessments or correlations with other measures.

The example provided is a multistep process that is utilized at one institution; however, it is important to note that one part of this process may be sufficient in other contexts. For example, a smaller community college may find it sufficient to provide faculty review and documentation of their assessment alignment with instructional objectives/student learning outcomes, without the formal review of higher-level committees. City Tech takes this more basic approach to establishing validity and provides *Content Validity* forms for the faculty to complete. The assignment and faculty documentation (*Content Validity* forms) are reviewed within a 90-minute Content Validity session that is scheduled within the assessment cycle.

The Role of Faculty in Establishing Reliability

Establishing evidence of reliability is crucial to guarantee the efficient use and interpretation of assessments in higher education. According to Cumming and Miller (2017), reliability is the consistency of scores assigned; this applies to multiple types of assessments, such as tests and performance appraisals. Thus, it is important to assess reliability when multiple raters are used in scoring assessments, when multiple items are included in an assessment, or when multiple forms of an assessment are used. The examples provide multiple sources of error variation in the assigned scores that need to be evaluated and subsequently minimized. Reliability studies enable the assessor and the assessed to have more confidence in the assessment procedure, and are a necessary requirement in establishing the validity of the instrument.

Q2

Reliability includes a broad range of methods based on both theoretical and statistical approaches. Methods for determining reliability require knowledge, support, time, and adequate software/analysis tools. When the assessment is locally developed at a university or college, the reliability could be estimated based on the administration(s) of the assessment and data analyses that range from percent agreement in scoring to correlational methods and analysis of variance to more complex modeling using item response theory. Depending upon the stakes attached to the assessment, the robust methodology should be considered. For example, if an institution requires its students to pass an exam to advance to junior level standing,

a more robust method should be utilized since the exam would be considered high-stakes. Each of the methods would require some familiarity with statistical methods. However, the statistical procedures can be as simple as estimating percent agreement or correlations in Excel, or estimating internal consistency estimates in R, SAS, or other software.

One frequently used method of examining reliability in higher education is to examine the consistency of scorers (faculty) for performance appraisal tasks. This can be useful when utilizing rubrics to score assessments or open-ended assignments such as capstone projects, speeches, and term papers. There are multiple examples of how universities have implemented the evaluation of inter-rater reliability.

Establishing Reliability for Program and General Education Assessments at City Tech

The Office of Assessment, Institutional Research and Effectiveness (AIRE) at City Tech has organized and facilitated inter-rater reliability sessions on a cyclical basis for various assessment efforts, including program assessments and general education/institutional outcomes assessment. The assessment system has embedded and documented these sessions within the assessment cycle. These inter-rater reliability sessions are also included on the assessment calendar so all faculty involved are aware of the event. Prior to the scoring session, faculty involved in the data collection, that is the faculty teaching the courses that were sampled for the data collection, are asked to submit samples of student work to the AIRE office that include exemplars classified as "excellent," "average," and "poor." These student artifacts are then de-identified, numbered, and archived for scoring within the inter-rater reliability session. This sample of student artifacts is then used to conduct the inter-rater reliability session by asking multiple faculty to score a subset of the artifacts. The artifacts are spiraled amongst the different raters and the raters are provided with the assigned task that was given to the students, the rubric and the artifact specimen set. Each faculty member (rater) will score the student work and leave their documentation with the AIRE office. It is important to note that the faculty who conducted the original assessment in their course are included in these sessions. The faculty who submitted assessment materials do not rerate their students' artifacts. The specimen sets are spiraled in such a manner that each artifact receives two additional ratings by at least two faculty. These sessions generally take 2 hours and are conducted once within the assessment cycle (e.g., once every 3 years for a 3-year assessment cycle).

The AIRE office maintains a database of the scores, artifact identification number, and the multiple scores by the different faculty. Then, an inter-rater reliability coefficient, ranging from 0 to 1, is computed for each performance indicator. This coefficient indicates the strength of the inter-rater reliability or consistency of ratings among the faculty. A coefficient of 1

2 indicates a perfect consistency among raters. In essence, if a student artifact
3 is assigned a low score for a particular performance indicator by one faculty
4 member on a clearly defined rubric, other faculty members also rated the
5 artifact as low for that same performance indicator.
6 　　After the results are evaluated, inconsistencies in rating student work
7 are noted. Faculty may convene to review the assessments and/or rubrics
8 to ensure clarity. After the inter-rater reliability session, the assessments
9 are reviewed and revised if necessary. Inter-rater reliability estimates have
10 ranged from 0.5 to 0.9. In the cases where the reliability coefficient was low
11 (e.g., 0.5), faculty typically made adjustments to the rubric to ensure clarity
12 of the criteria for score assignment.
13 　　The University of Florida engages in a more sophisticated reliability
14 process and builds in various statistical analyses for various levels of assess-
15 ments. Although the assessment office at the university is relatively small for
16 the size of the institution, it capitalizes on the expertise available through its
17 educational measurement and psychometrics department faculty and con-
18 ducts robust reliability analyses.

The Role of Faculty in Reviewing Assessments for Fairness

22 Fairness is the third foundational section of *The Standards* and could be con-
23 sidered as a part of the validity argument for an assessment. Fairness can be
24 defined as having equivalent uses and interpretations across *all* examinees,
25 that is, the interpretations and uses of the assessment are equitable regard-
26 less of a person's race, ethnicity, gender or sexual orientation, disability, or
27 any other relevant group. Fairness should be considered and documented
28 at all stages of assessment from development through final uses and inter-
29 pretations.
30 　　Fairness is especially important in the higher education environment
31 that focuses on diversity and multicultural perspectives. At the University of
32 Florida, the Mission Statement includes a call to "create the broadly diverse
33 environment necessary to foster multicultural skills and perspectives." A
34 key component of UF's SACSCOC accreditation process was the emphasis
35 on internationalization through their Quality Enhancement Plan. City Tech,
36 which is consistently one of the highest ranked institutions with respect to
37 diversity by the U.S. News and World Report, also has a strong emphasis on
38 diversity and inclusion. Its mission emphasizes learning in a "diverse urban
39 population" and its goals are based on "diverse perspectives." The faculty
40 at City Tech are keenly aware of the importance of avoiding biases.
41 　　The recognition of increasing diversity within the nation is a docu-
42 mented trend, where studies project that the white population will be a
43 minority in the United States by 2045 (Frey, 2018). This demographic shift
44 necessitates fairness reviews. There are many methods for documenting fair-
45 ness with differing assumptions and issues. Two of the most widely used
46 methods for examining fairness at colleges and universities are: (1) faculty

2 review of items for offensiveness or sensitivity of the content, and (2) statis-
3 tically determining if there are systematic differences in performance on an
4 item across subpopulations after controlling for overall ability. Depending
5 upon the institution's resources, both of these methods may be considered
6 or perhaps just one.

7 Offensiveness or sensitivity reviews can be conducted with any assess-
8 ment. As mentioned above, faculty can review the content of the assessment
9 and the expected responses to determine that they would be consistent and
10 fair across different groups. Experts (i.e., faculty) simply respond to ques-
11 tions such as, "Reviewing this item, is it offensive or insensitive to (subpop-
12 ulation)?" The faculty review is not required, yet strongly encouraged at the
13 University of Florida. At City Tech, fairness is reviewed by the faculty and a
14 judgment or recommendation is made during the Content Validity sessions
15 that are conducted within the assessment cycle.

16 Universities that wish to engage in a more advanced fairness review
17 may employ empirical techniques by examining the assessments at test,
18 item, or task level. While an item or task may or may not use offensive
19 language, actual differences in performance should be examined when pos-
20 sible. Fairness is potentially assessed through several statistical indices mea-
21 suring Differential Item Functioning (DIF). DIF analyses can be complex
22 and may require a consultant or the participation of faculty with the needed
23 statistical expertise. The University of Florida, capitalizing on faculty exper-
24 tise within its educational measurement and psychometrics program, con-
25 ducted such a DIF analyses for the Quality Enhancement Plan, as it was a
26 university-wide program. However, this institution is more advanced than
27 most in its assessment methodologies.

28
29
30 ## Improvements Based on Assessment Results

31 Assessment efforts at City Tech have been a meaningful and effective part
32 of the teaching and learning processes. When data are available to better
33 understand student weaknesses, faculty can work collaboratively to identify
34 effective strategies to improve student learning. City Tech faculty utilized
35 assessment data from a college-wide assessment of reading skills, using the
36 AAC&U LEAP VALUE assessment tool. Results of this large-scale assess-
37 ment across the college indicated that the majority of students sampled
38 did not meet faculty criteria for reading, confirming faculty assertions that
39 their students struggled with textbook readings, as well as readings assigned
40 within their course work.

41 Based on the assessment results, City Tech initiated *Reading Effectively*
42 *Across the Disciplines* (READ), a college-wide program developed to address
43 reading deficiencies (Cumming, Deiner, & August, 2017). A comparable
44 sample of students was assessed during the next assessment cycle. An eval-
45 uation of the results indicated a significant improvement: a majority of the
46 students met or exceeded the faculty target of proficiency. On the other

hand, the admission indices were not significantly different ($p > 0.05$) between the two cohorts of students. Therefore, this improvement was not attributed to differences in the two cohorts. While City Tech does not purport that the READ program is the sole cause of the increase in reading skills, it does acknowledge the value of assessment data in determining the needs of its students and launching meaningful improvement strategies.

How Assessment Efforts Have Led to Improved Retention Rates

In 2009, a formal assessment system was introduced at City Tech. Since then, the College has realized an increase in the number of degree completers, as well as retention rates in programs where the assessment process is highly valued. For example, a 17% retention rate increase over a 6-year period was realized for a program in the School of Professional Studies. In this department, faculty were tasked with identifying a critical course to monitor on a department level; they opted to observe a course that required mandatory first-time completion for continuation within the degree program. It was therefore hypothesized that the improvement of student outcomes within this course would have a positive impact on overall retention rates, and by extension, graduation rates. Since this program also requires professional certification, we are confident that the improvement is not a result of grade inflation or lowering the expectations of the students since the student pass-rates on the professional exams are extremely high (more than 90% of the program's students become certified).

The purpose of these examples is to emphasize that assessment can be effective when used properly; in the cases presented, assessment implementation was highly dependent upon faculty leadership and the utilization of a respected assessment system that enabled faculty to be convinced of its resulting data quality. The most effective way to obtain this respect by the faculty is to present them with an assessment system that builds upon the three core principles of educational measurement: validity, reliability, and fairness.

Communicating With Accreditation Authorities During the Site Visit

It is important that college administrators communicate with evaluation team reviewers about the use of assessment methods. It is also important to ensure that all documentation that has been submitted to accrediting teams with respect to student learning outcomes assessment is aligned and documented with the program, college, and university mission and strategic plans. This holds true for administrative, educational, and student support unit assessments.

Inevitably, evaluators make requests for additional documentation when they are on their site visit. Following is a list of suggested assessment

documents colleges should have available for reviewers during accreditation self-study visit:

Strategic Plan for the Assessment Office/Committee/Governing Unit
- Mission, goals, and outcomes
- Long-term plans (3–5 years)
- Key milestones and deliverable dates
- Roster of faculty and staff who contributed to drafting the plan
- Description of the plan approval process

Organizational Structure of the Assessment System (Including Committee Structures)
- Organizational charts
- Committee structures
- Rosters of committee membership

Assessment Policies
- Assessment cycle length
- Self-study schedule (maximum length)
- FERPA issues
- Assessment data usage for research and publication
- Data retention policy

Curriculum Maps for Each Degree Program
- Program-level student learning outcomes
- Courses required for degree program (required) and elective courses (optional)
- Alignment indicated (sometimes with the level of contribution, e.g., Introduced, Reinforced, etc.)

Library of Assessment Meeting Documentation
- Agendas
- Minutes
- Attendance records

Assessment Report of Accomplishments for Each Academic Year
- Provide an overview of highlights of accomplishments (e.g., created a General Education Assessment Brief series for distribution to all constituencies)
- Demonstration of upgrades within the office (e.g., dashboards)
- Conducting workshops, indicating the number of workshops, topics, and number of faculty/staff in attendance

Assessment Cycle Calendar—Programs (and Courses, If Required by Accreditation Governing Body) and General Education/Institutional Outcomes
- Detail the time/semester for data collection for each student learning outcome
- Indicate the faculty/staff responsible for overseeing the data collection for each SLO
- Detail the time/semester for meetings to discuss the results

○ Detail the time/semester for meetings to develop improvement strategies for any SLOs requiring action

Assessment Validation Processes and Results
○ Description of validation procedure
○ Roster of faculty involved in establishing/reviewing validity
○ Validation worksheets and results

Assessment Reliability Processes and Results
○ Description of reliability procedure
○ Roster of faculty involved in inter-rater reliability meetings (if applicable)
○ Reliability worksheets and results

Assessment of Fairness Processes and Results
○ Description of fairness procedure
○ Roster of faculty involved in fairness review
○ Fairness worksheets and results

Assessment Database
○ Courses
○ Instructors
○ Number of students assessed
○ Types of assessment (program, general education, course)
○ Time of assessment

Assessment Instruments
○ Rubrics
○ Performance appraisals, tests
○ Student exemplars
○ Description of how the assessments were developed, when, and by whom

Library of the Assessment Reports
○ Reports submitted by each program on a cyclical basis (e.g., annual reports)—see Table 11.1 for suggested elements of the reports
○ Summary documentation of the improvement strategies from the cyclical reports

  • Improvement strategies for each applicable student learning outcome

Evaluation of the Quality of the Assessment System (a Newer Standard Within Some Regional Accreditation Bodies)
○ Description of the internal structure to review the quality of assessment reports submitted, adequacy of sampling design, description of the procedures to ensure validity, reliability, and fairness is addressed, and meaningful improvement strategies are implemented as necessary
○ Description of any external review of the assessment system and quality of assessment reports
○ Qualifications of the reviewers

Table 11.1   Suggested Elements for Inclusion With the Program
Assessment Report

The program enrollment (number of students enrolled in the program).
The list of student learning outcomes (which should be published in the college
    catalog and listed on the website).
The length of the assessment cycle (e.g., 3 years).
An overview of the assessment timeline.
The course(s) selected to assess each program's student learning outcomes.
A description or copy of the assessment instrument(s) and the list of faculty involved
    in the instrument's development, including *when* the assessment was developed/
    revised.
Sampling methods of the courses (e.g., number of sections, number of students
    assessed), as well as the faculty who participated in the data collection and scoring.
A description of how the assessment was scored.
The faculty target for success.
The results of the assessment presented in a table and/or graphic format.
An evaluation summary from faculty discussions of the results, including when the
    meeting(s) were held and participating faculty.
Improvement strategies that faculty identified, and discussion of how they will be
    implemented.
A scheduled reassessment to evaluate the effectiveness of the improvement strategies.
Content validity documentation.
Inter-rater reliability and/or internal consistency, or other form of reliability
    documentation.
Fairness documentation.

Providing relevant and useful assessment documentation is essential for a successful accreditation visit. Faculty should be provided with a list of elements that are required for their assessment reports and provided with exemplars from within the institution. Table 11.1 provides a list of elements that should satisfy both, regional and professional, accreditation governing bodies.

In addition to having documentation available, it is advisable to prepare faculty and staff for the in-person meetings with the evaluation team to discuss any questions that may arise from the review of the documentation submitted or any omissions they wish to discuss. It is also necessary to identify a team of faculty, students, and staff who can be "on call" for each program and who can be available during the site visit. This on-call team should be advised that they may be asked to answer important assessment questions for their program that include the entire assessment process from development through uses for program improvement. All faculty, students, and staff that are identified to serve in this capacity should be well versed with respect to the documentation previously submitted to the evaluation team. A preparation meeting in advance of the site visit is advisable. Some faculty, students, and staff can become anxious at the thought of being called for a meeting with the evaluation team. Having these meetings to inform them of the likely agenda with the evaluation team can prove to be highly beneficial.

Particularly, it is helpful if each of the "on-call" team members can speak to the particular improvements that have benefited their programs as a result of the assessment activities and how their program, in general, is working to attain the mission and strategic plan goals at the various levels.

References

AERA (American Educational Research Association), APA (American Psychological Association), and NCME (National Council on Measurement in Education). (2014). *Standards for educational and psychological tests and manuals*. Washington, DC: American Educational Research Association.

Cumming, T., Deiner, L. J., & August, B. (2017). Case study: The New York City College of Technology approach to general education assessment. In T. Cumming & M. D. Miller (Eds.), *Enhancing assessment in higher education: Putting psychometrics to work* (pp. 151–172). Sterling, VA: Stylus.

Cumming, T., & Zhao, Y. (2015). Post-secondary regional accreditation assessment standards in the United States. *European Scientific Journal, Special edition Vol. 1*, 363.

Ewell, P., & Cumming, T. (2017). History and conceptual basis of assessment in higher education. In T. Cumming & M. D. Miller (Eds.), *Enhancing assessment in higher education: Putting psychometrics to work* (pp. 3–26). Sterling, VA: Stylus.

Frey, W. (2018). *The US will become 'minority white' in 2045, Census projects: Youthful minorities are the engine of future growth*. Retrieved from https://www.brookings.edu/blog/the-avenue/2018/03/14/the-us-will-become-minority-white-in-2045-census-projects/

Linn, R. (2003). Accountability: Responsibility and reasonable expectations. *Educational Researcher*, *32*(7), 3–13.

Miller, M. D., Linn, R., & Gronlund, N. M. (2013). *Measurement and assessment in teaching* (11th ed.). Boston, MA: Pearson.

Rice, J. (1914). *Scientific management in education*. New York: Hinds, Noble & Eldredge.

US Congress Office of Technology Assessment. (1992). *Testing in American schools: Asking the right questions, OTA-SET-519*. Washington, DC: US Government Printing Office.

US Department of Education, National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform*. Washington, DC: U.S. Government Printing Office.

Zwick, R. (2012). *A review of ETS differential item functioning assessment procedures: Flagging rules, minimum sample size requirements, and criterion refinement* (Report No. RR-12-08). Princeton, NJ: ETS.

Q4

TAMMIE CUMMING is director of the Office of Assessment, Institutional Research and Effectiveness at New York City College of Technology (CUNY).

M. DAVID MILLER is a professor of research methods and evaluation at the University of Florida.