Fairness and Equity in Assessment: Challenges and Opportunities

Maria Cristina Leite, EdD
Assistant Director, Institutional Assessment, Office of the Provost, University of Florida, Gainesville, Florida USA

Timothy S. Brophy, PhD
Professor and Director, Institutional Assessment, Office of the Provost, University of Florida, Gainesville, Florida USA

Today's Goals

Examine the concepts of fairness and equity as central to assessment practices in higher education

Explore opportunities to develop guidelines as tools to improve the assessment process

Share models of fairness and equity in assessment in higher education

Discuss opportunities for development and application of fairness and equity in assessment guidelines in your institution

Meeting format

- **Time:** We will work together for 40 minutes, then take a 10-minute break, and reconvene

- **Breaks:** There will be two 10-minute breaks during the workshop, which will be determined at reasonable stopping times

- **Group formation:** We will alternate between whole-group information sessions and breakout group discussions – there are five breakout sessions

- **Breakout Groups:**
  - There are three pre-assigned breakout groups. You'll remain in the same breakout group for the entire workshop.
  - One person in the group will need to self-select as the leader who will summarize the group's responses to the discussion questions.
  - Breakout sessions are 5 minutes each; breakout group reports are 2 minutes each.

Virtual Polling used this session

- We will use virtual polling to ask questions and to collate audience responses

- To access and contribute to our presentation polls, please go to

Pollev.com/marialeite275 on your phone or computer

OR

Download the Poll Everywhere app on your phone and join the presentation

I was able to access Poll Everywhere

Getting to Know You

To respond to these polling questions, please go to PollEv.com/marialeite275

Where is your institution located? Please enter the name of the state. If you are outside the US, please enter the country only.

How many students are enrolled in your institution?

What is the highest degree your institution awards?

Why is fairness and equity in assessment important?

Part 1: Examining the concepts of fairness and equity as central to assessment practices in higher education

Fairness

- "Fairness is not viewed as something added on at the end of an assessment, but instead permeates throughout the entire process" (Cumming and Miller, 2017, p. 40)

- "Validity and fairness are closely connected." Assessments that "systematically misrepresent the standing of some individuals or some groups of individuals on the construct being measured" are unfair. Therefore, interpretation and use of results are not valid (Jankowski & Lundquist, 2022, p. 45).

- [the presence of] validity [evidence] of score interpretations for intended use(s) for individuals from all relevant subgroups. A test that is fair minimizes the construct- irrelevant variance associated with individual characteristics and testing contexts that otherwise would compromise the validity of scores for some individuals. (Standards for Education and Psychological Testing, 2014, p. 219).

Equity

- "Equity means acting to remedy injustices" (Montenegro & Henning, 2022, p. 5).

- Equity involves transparency in design (Jankowski & Lundquist, 2022, p. 38)

- "Equity in assessment is also referred to as culturally responsive assessment, culturally relevant assessment, inclusive assessment, and equitable assessment (Bevitt, 2015; Kaur et al., 2017; Montenegro & Jankowski, 2017a; Singer-Freeman et al., 2019; Slee, 2010; Suskie, 2000, as cited in Henning, et al., 2022, p. 146)

Defining Fairness and Equity in Assessment

- Fairness and equity in assessment ensure that no person is disadvantaged based on individual characteristics so that all have unobstructed opportunities to demonstrate their standing on the construct or criterion being measured, and that full access to assessments and the results of assessments are guaranteed.

Fairness and Equity in Context

- **Fairness** is a fundamental validity issue
  - How assessments are sensitive to various subgroups among our populations of students is a foundational concern (e.g., individuals with disabilities, those with limited English proficiency, students from other cultures or countries)

- **Equity** – two concepts to bear in mind:
  - *Accessibility* – all should have an unobstructed opportunity to demonstrate their standing on the construct(s) being measured
  - *Universal design* – the design of the assessment maximizes accessibility for all

Evolution of Assessment

Assessment of Student Learning

- Measurement & Testing
- Focus on summative final grades

Assessment as Part of Program Review

- Summative piece determines program effectiveness

Assessment for Improvement

- Student-centered
- Consequential
- Meaningful
- Valid use and interpretation
- Mindful of "(in) equity"

Jankowski, N. A. (Ed.), Baker, G. R. (Ed.), Montenegro, E., & Brown-Tess, K. (2020). *Student-focused learning and assessment: involving students in the learning process in higher education*. Peter Lang Publishing. https://doi.org/10.3726/b16909

Equitable and Fair Assessment

In your experience,

- What are effective ways higher education institutions can ensure equitable and fair assessments?

Your Institution

- Identify potential opportunities for implementation of fairness and equity in assessment in your institution.

Why does it matter?

Part 2:
Exploring opportunities to develop guidelines as tools to improve the assessment process

Opportunities for Improvement

Summer 2020

- "We can make progress - in education, in advancing truth, reconciliation and justice, and in anti-racism, equality and working to eradicate inequities." (Fuchs, K., June 18, 2020)

Fall 2020

- Dr. Timothy Brophy, Director of Institutional Assessment:
  - formed a broadly representative Task Force to lead the development of guidelines to address fairness and equity in assessment at the University of Florida
  - Over 50 nominations (Faculty, staff, student)
  - 33 accepted
  - Co-chairs:
    - Dr. Corinne Huggins-Manley, College of Education;
    - Dr. M. David Miller, College of Education; and
    - Dr. Teresa Mutahi, College of Liberal Arts and Sciences.

The Charge to the Task Force

Establish a set of guidelines and models for UF faculty, instructors, staff, and administrators to help ensure fairness and equity in assessment in all contexts at the university. The guidelines must address fairness and equity in the entire assessment process, including development, opportunity to learn, administration, measurement, interpretations and uses of the results, and the evaluation of the measurement properties of the assessment.

Shared Understandings and Beliefs

Shared Understandings
- Assessment is integral to advancing the mission of the university
- Guidelines are not absolutes
- Guidelines are applied in inclusive and productive ways

Beliefs
- Assessment is a valuable process in higher education
- Fairness and equity are inherently important to the assessment process
- Assessment is most effective when implemented fairly and equitably

Task Force

How did the work evolve?

Groups:

1. Guidelines
2. Models

1. **Guideline Development Group**

• Develop the guidelines and review the recommendations from the model development group, and where mutually agreeable, implement the recommendations to modify the guidelines.

2. **Model Development Group**

• Locate and develop models that operationalize the guidelines in various contexts; provide the guideline development group with modifications to increase their utility.

Stages of Assessment

- Assessment Development

- Assessment Selection

- Assessment Administration

- Opportunity to Learn (preparation for the assessment)

- Assessment Scoring

- Assessment Score Interpretations and Uses

- Evaluations of Assessments (e.g., validity and reliability evidence)

Four Primary Groups

Work in progress with focus on the following groups:

- Student

- Faculty

- Staff

- Applicants

Student Assessment

For the purpose of this workshop, we will focus on fairness and equity on student assessment.

Student Assessment

How are students assessed at UF?

Examples:

Course-based assessments (e.g., assignments, papers, portfolios)

Scholarship and award applications

Program-level assessments (e.g., qualifying exams, clinical assessments, practica/internship/assistant ship assessments, capstone)

University-level assessments (e.g., graduation survey, academic spoken English)

Guidelines

Fairness and equity in the entire assessment process, including:

Development

Opportunity to learn

Administration

Measurements

Interpretations and uses of results

Evaluation of the measurement properties of the assessment

Guidelines are intended to:

Support analysis of fairness and equity in assessment

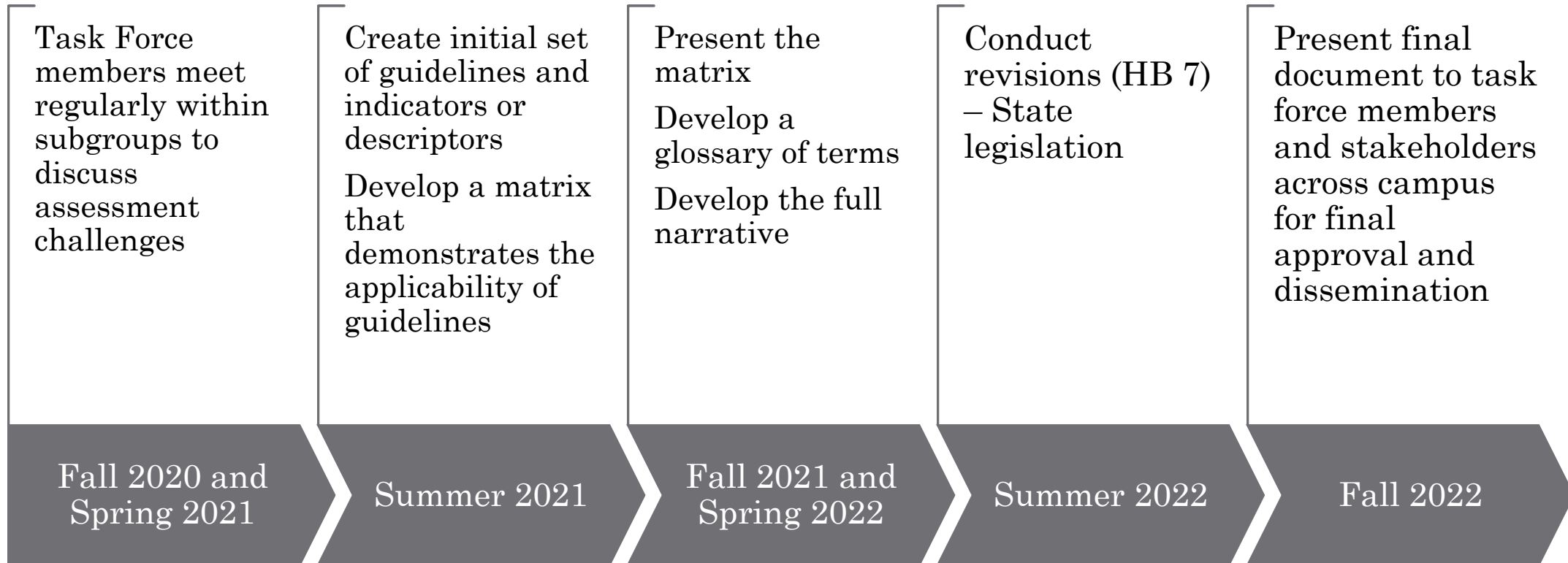Strengthen fairness and equity in assessment

Contextualize assessment practices

Breakout Group Discussions – 5 minutes

Considering your institution's context, discuss with your peers the opportunities you envision to implement, continue, and/or effectively promote fairness and equity in student assessment.

The leaders of each group will share their main discussion points for 2 minutes after we reconvene.

Task Force Preliminary Work

Task Force members meet regularly within subgroups to discuss assessment challenges	Create initial set of guidelines and indicators or descriptors Develop a matrix that demonstrates the applicability of guidelines	Present the matrix Develop a glossary of terms Develop the full narrative	Conduct revisions (HB 7) – State legislation	Present final document to task force members and stakeholders across campus for final approval and dissemination
Fall 2020 and Spring 2021	**Summer 2021**	**Fall 2021 and Spring 2022**	**Summer 2022**	**Fall 2022**

The Guidelines

Compiling the taskforce work into guidelines, indicators, and models

Translating Taskforce Notes

Guideline	Students	Faculty	Staff	Applicants
Guideline 1 - Assessment Development				
1.1 *Assessment developers (e.g., faculty, teaching assistants, etc.) should evaluate content and tasks for the degree to which their results support the decisions and uses being made from the assessment. The assessment development process should reflect:*				
1.1.a. Clearly defined constructs, and guidelines for ratings (if applicable), interpretations, and recommendation uses.	x	x	x	x
1.1.b. For course assessments, a clear alignment of assessment content with Student Learning Objectives of the course. In the case of a Research course or internship, clear guidelines for evaluations from the start of the semester.	x			

Current Matrix

Improvements based on feedback

Guidelines (Student Assessment)
Guideline 1 - Assessment Development
1.1 Assessment developers (e.g., faculty, teaching assistants, etc.) should evaluate content and tasks for the degree to which their results support the decisions and uses being made from the assessment.
1.1.a. Assessment reflects clearly defined constructs, and guidelines for ratings (if applicable), interpretations, and recommendation uses.
1.1.b. For course assessments, there is a clear alignment of assessment content with Student Learning Objectives of the course. In the case of a Research course or internship, there are clear guidelines for evaluations from the start of the semester.
1.1.c. For program assessments, there is a clear alignment of assessment content with the program student learning outcome for the course.
1.1.d. Assessments are flexible and adaptable to meet cultural characteristics, diverse identities, and ways of knowing (i.e., Culturally responsive).
1.1.e. Assessment development is a collaborative process that includes the voices of those who will be assessed (e.g., student engagement).

Breakout Group Discussions – 5 minutes

Examine the set of guidelines provided and reflect on those that could be adapted for your institution context.

The leaders of each group will share their main discussion points for 2 minutes after we reconvene.

Scenario: Fairness and Equity

- In response to CDC guidance to control the spread of COVID-19 the University of Lake Wobegon has suddenly gone fully online during week 5 of the 14-week semester and sent students home to complete the semester at a distance. The university requires that students own a laptop with a minimum set of specifications, but in reality many students rely on the computers in the library or other student learning centers on campus to access course material online.

- For their final examination in architecture, students must develop a building plan and blueprint. The assessment assumes that students have access to stable internet, appropriate software, and computer resources to complete the exam.

- What would you advise the faculty regarding fairness and equity?

Breakout Group Discussions – 5 minutes

How would you determine the fairness and equity of assessment in these situations? The leaders of each group will share their main discussion points for 2 minutes after we reconvene.

Two approaches that can strengthen assessment of student learning

Part 3: Sharing models of fairness and equity in assessment in higher education

Consensus-Based Assessment

What it is, how it works

The Premise

Data collected as evidence of learning is delimited by assessment type

Consensus – based methods present a non-delimited approach to assessment

The Quandary

The assessment of creative or open-ended works is difficult to standardize

There is no 'standardized student' nor is there a 'standardized response' to an open-ended or creative assessment task

Even our best attempts at developing standardized rubrics or scoring approaches are limited in their scope, transferability to practice, and still raise questions of validity

If variability is systematically ignored, individuals become synonymous with statistical averages, and faculty and researchers lose the ability to account for the very processes that underpin the phenomena they seek to explain

The "Science of the Individual"

Rose et al. (2013) present an approach to understanding and analyzing human behavior based on the precept that individuals behave, learn, and develop in distinctive ways, showing patterns of variability that are not captured by models based on statistical averages

The authors ask us to consider human beings as *dynamic systems*, which assumes that behavior is actively organized and context-dependent, variability is expected as a natural outcome

They argue that learning is not a linear progression through a universal sequence, where the start and endpoint are predetermined

Creative and open-ended tasks and their responses often fall into this category

Individualized Responses

Reconsidering our approach to assessment

Based on our understanding of the individual variability of creative human responses, we should reconsider our approach to assessment data collection and analysis

Standard assessment methodology: aggregate, then analyze

A *reconsidered* methodology: Analyze, then aggregate

Why consensus-based assessments?

- Most assessments are designed to be measured by a predetermined set of criteria using a tool such as a rubric or checklist, or by counting points earned by responding to questions (such as those on an exam or quiz) and placing the total score on a grading scale to determine achievement.

- However, there are assessment types for which the development of preset criteria is counterproductive because the criteria *constrain* the response by forcing respondents to *conform* to the preset criteria.

Why consensus-based assessments?

- These assessments include those that examine individual distinctiveness or creativity and therefore result in considerable expected response variability – including, but not limited to: interpretation, creative writing (stories, novels, poems, etc.), artistic creations (musical compositions, improvisations, choreography, paintings, sculptures, ceramics, and other fine arts), development of new theories, logical arguments, etc.

- Setting criteria in advance for these types of assessments could lead to response conformity that conflicts with their intended purpose to express the individuality of the respondent.

The Concept of *Consensual Assessment* (to measure creativity)

The concept of consensual assessment (Amabile, 1996) rests on the belief that validity evidence for assessments of creativity is strongest when experts rate the creative product using their subjective judgments

In the consensual assessment process, raters use predetermined criteria (dimensions) as a focal point for measurement, and determine levels of achievement using some form of scale that leaves determination of levels of criterion demonstration to the raters' subjective judgments.

Consensual Assessment

Judges measure the creative products of interest in isolation – there is no collaboration or contact among the judges

Interrater reliability is critical and has been shown to be in acceptable ranges in some recent studies (Hennessey et al., 2011), but not all (Hickey, 2001).

Consensus Moderation

Rethinking our approach to the measurement of performance-based, open-ended, and creative works to accommodate individual variability

Consensus Moderation Defined

Consensus: reaching a general or common understanding

Moderation: by definition, lessening of extremes.

Moderation can be done in different ways: (1) *Averaging* different readings or coded judgments; (2) *removing* the most discordant judgment or divergent judgments (and averaging the remainder); (3) Accepting the *middle* reading; (4) *Discussing* until **consensus** is reached.

Consensus moderation is a *result* of successful consensus-seeking which reduces discord and by that means, moderates.

Consensus Moderation as an assessment process

Multiple 'experts' engaged in the review of an artifact or work

Assessors must be open to the qualities observed in the work; there is no attempt to steer the student toward any particular qualities

Assessors make a holistic judgement about the level of proficiency or competence

Students need to be inducted into this process – so they can learn to monitor and control the quality of their own performances and productions while their productions are in progress.

Reliability and Validity

Consensus-moderation provides strong validity evidence in that there is discussion among the assessors to arrive at consensus on the assessment decision

Interrater reliability is high when all assessors reach consensus

Breakout Group Discussions – 5 minutes

How would you address these situations at your institution? The leaders of each group will share their main discussion points for 2 minutes after we reconvene.

Consensus-based Approaches in Context

- At your institution, in which programs might assessment of student work be best approached using consensual assessment or consensus moderation?

- How might you advise faculty to triangulate quantitative assessment data from exams, quizzes, etc. with the results of a consensus-based assessments?

Building a Collaborative Rubric

Guiding faculty to identify criteria and levels of achievement

Our example: Performance/Production

- A faculty member has developed a new general education course, *Russian Fairy Tales*.

- The course description:

Be they manifested in stories, fiction, film, music, popular culture, or everyday life, folk tales and beliefs play an outsized role in shaping individual, social, and national identity. "Russian Fairy Tales" delves into the rich tradition of Slavic lore, employing a variety of critical tools and cross-cultural perspectives (the Grimms, Perrault, Disney, etc.) to better understand and appreciate the magic of their enduring impact.

Our example: Performance/Production

- For the assessment, students can demonstrate their learning in a multiple ways. One of the ways is through a performance or production. A *performance/production* is a literary (e.g., story, poem, play, libretto, essay, critique) or artistic work (music, dance, drama, visual art, media), presented or exhibited to the public on stage, screen, or in a physical or digital space. These are open-ended, creative works.

- We will work together to help the faculty member to develop a rubric for this course.

Basic Steps for Leading Faculty to Develop Rubrics for Hard to Measure Disciplines

- Guide the faculty member to develop the criteria for this rubric.

- Determine how many levels of achievement are appropriate

- Use nominal descriptors for each level

- Write complete descriptions of each level of the rubric

Breakout Group Discussions – 5 minutes

What would you do to guide this faculty member to develop criteria, levels of achievement, and writing descriptions of each achievement level for this rubric? The leaders of each group will share their main discussion points for 2 minutes after we reconvene.

If we have time – an example

This example was developed by the University of Florida Quest Assessment Task Force

Performance/Production Rubric

|	Criteria	4	3 (Target)	2	1
Quest 1 Outcomes — **Content**	**Components** *Uses appropriate components to evoke and elevate the audience experience.*	The performance/production intentionally, skillfully and knowledgably integrates well-chosen, appropriate components that lead to a compelling audience experience.	The performance/production knowledgeably integrates specific components that lead to the intended audience experience.	The performance/production shows knowledge of performance/production components, but inconsistently integrates and implements the components, diminishing the effectiveness of the intended audience experience.	The performance/production shows little knowledge of performance/production components. Specific components necessary to evoke the intended audience experience are not effectively integrated or implemented. The effectiveness of the performance/production is lost or significantly diminished.
Critical Thinking	**Creativity** *Displays innovative or transformational ideas, formats, or solutions about the topic.*	The performance/production displays innovative and transformational ideas, formats, or solutions about the topic that lead to a compelling and effective audience experience.	The performance/production displays innovative and transformational ideas, formats, or solutions about the topic that lead to an effective audience experience.	The performance/production uses conventional ideas, format, or solutions about the topic, diminishing the audience experience.	The performance/production uses borrowed or unoriginal ideas, formats, and solutions about the topic, leading to a significantly diminished audience experience.
Communication	**Interpretation** *Conveys intended meaning and purpose.*	The performance/production is focused, coherent, compelling, and effectively conveys the performer/producer's meaning and purpose.	The performance/production is focused, coherent, and conveys the performer/producer's meaning and purpose effectively.	The focus of the performance/production wavers and diminishes the performer/producer's meaning and purpose because of periodic incoherence.	The performance/production is incoherent and lacks focus. The performer/producer's meaning and purpose are lost or so significantly diminished that they are imperceptible.
Communication	**Delivery** *Uses media, tools, materials, and/or technology effectively to maximize delivery.*	The performance/production fully engages the audience using media, tools, materials, and technology appropriately and intentionally integrated to create a compelling audience experience.	The performance/production engages the audience using media, tools, materials, and technology to create an effective audience experience.	The performance/production engages the audience inconsistently. Media, tools, materials, and technology are used in a manner that diminishes the audience experience.	The performance/production fails to engage the audience. Media, tools, materials and technology are misused or so ineffective that the audience experience is significantly diminished.
Connection	**Reflection** *Connects the learning experience to the student's intellectual, personal, and/or professional development at UF and beyond.*	A thoughtful, profound, and insightful connection of the learning experience to previous learning and/or the student's intellectual, personal, and professional growth.	A basic connection of the student's learning experience to previous learning and/or the student's intellectual, personal, and professional growth.	A partial and incomplete connection of the student's learning experience to previous learning and/or the student's intellectual, personal, and professional growth.	Provides a superficial connection of the student's learning experience to previous learning and/or the student's intellectual, personal, and professional growth.

Questions

A Closing Thought

The greater danger for most of us lies not in setting our aim too high and falling short; but in setting our aim too low and achieving our mark.

-Michelangelo

Source: https://www.michelangelo.org/michelangelo-quotes.jsp
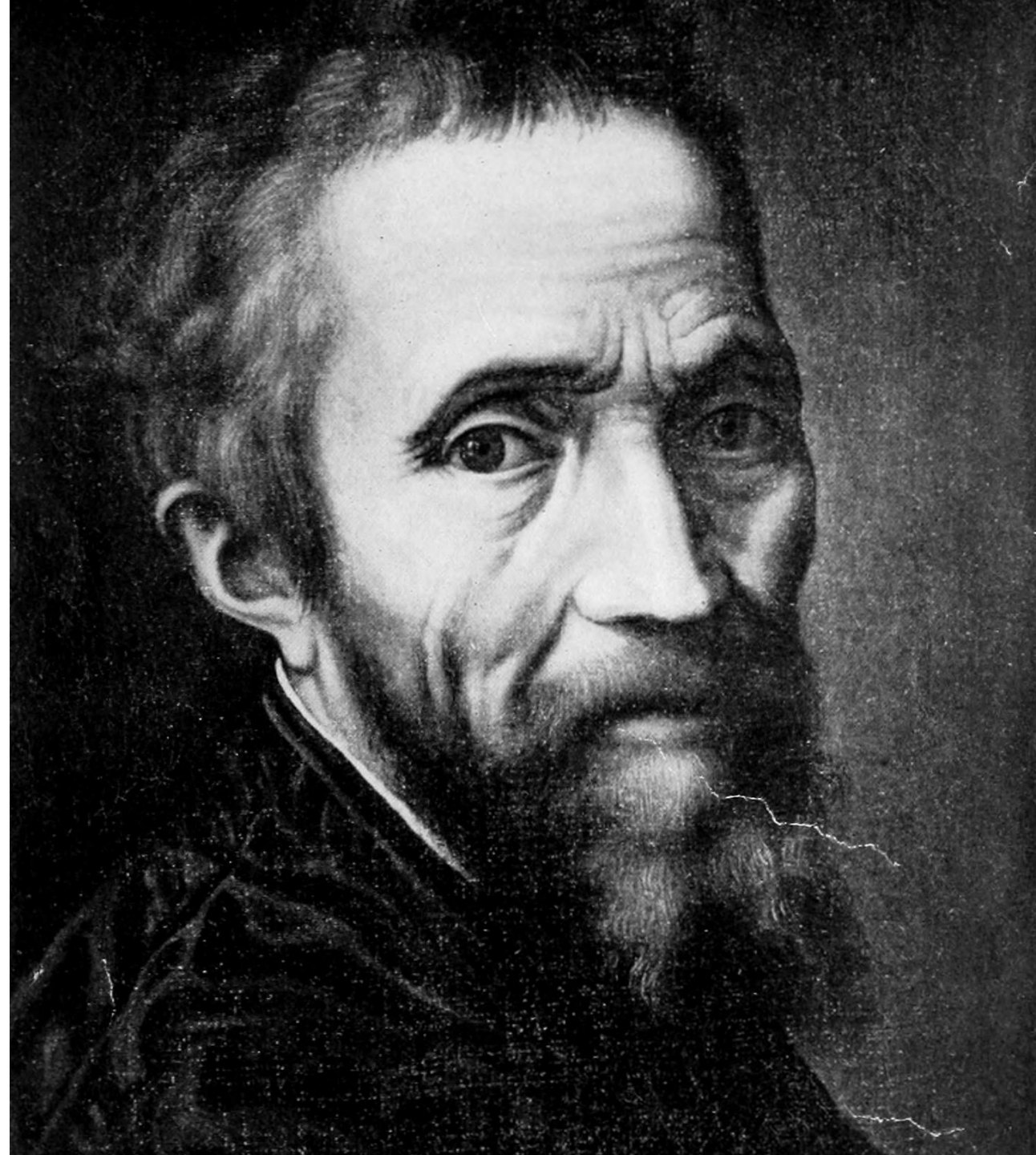
References

- Amabile, T. (1996). *Creativity in context: Update to the social psychology of creativity*. Boulder, CO: Westview Press.

- Hennessey, B. A.; Amabile, T.M.; Mueller, J.S.. (2011). Consensual assessment. In Runco, M.A. and Pritzker, S. R. (Eds.), *Encyclopedia of creativity* (2nd Ed., Vol. 1) (pp. 253-260). San Diego, CA: Academic Press. Retrieved from https://klm68f.media.zestyio.com/encyclopedia-of-creativity_9.pdf

- Hickey, M. (2001). An application of Amabile's consensual assessment technique for rating the creativity of children's musical compositions. *Journal of Research in Music Education, 49,* 234–244. doi:10.2307/3345709

- Rose, L. T.; Rouhani, P.; Fischer, K.W (2013). The science of the individual. *Mind, Brain, and Education 7*(1),152- 158

- Sadler, D. R.(2009, Spring). Are we shortchanging our students? The use of preset criteria in assessment. *TLA Interchange 3,* pp. 1-10. The University of Edinburgh: Edinburgh, Scotland. Retrieved from: http://blogs.ubc.ca/chendricks/files/2013/01/Sadler2009-AreWeShortchangingOurStudents.pdf

- Sadler, D. R. (2015, August 27). *Consensus moderation, backwards assessment, and calibration.* Assessment in Music Symposium, Queensland Conservatorium, Brisbane, Australia. Retrieved from http://assessmentinmusic.com.au/wp-content/uploads/2012/06/Royce-Sadlers-speaking-notes.docx

- University of Florida (2020). UF Quest Pilot Assessment Faculty Guide. Retrieved from https://assessment.aa.ufl.edu/media/assessmentaaufledu/academic-assessment/uf-quest-1-pilot-assessment/UF-Quest-Assessment-Faculty-Guide-and-Rubrics.pdf.

THANK YOU!

Timothy S. Brophy, Ph.D.

Professor and Director, Institutional Assessment

University of Florida, Gainesville, FL USA

Email: tbrophy@aa.ufl.edu

Office Phone: +1-352-273-4476

Maria Cristina Leite, Ed.D.

Assistant Director, Institutional Assessment

University of Florida, Gainesville, FL USA

Email: mleite1@ufl.edu

Office Phone: +1-352-392-2478