Center for Educational Measurement, Inc.

THE TEST OF EXCELLENCE

FORWARD
Focused Goals. Maximum Results.
2023-2028

46th ANNIVERSARY

VIRTUAL CONFERENCE

PROCEEDINGS

Building a Balanced Assessment System

BEYOND THE NUMBERS

30 SEP 2024 8:00 AM - 12:00 NN VIA zoom

www.cem-inc.org.ph

EDITORIAL TEAM

Chief Editors: *Armi S. Lantano*
Pristine Mae T. Cammayo

Editorial Advisers: *Dr. Grace H. Aguiling-Dalisay*
Iris Lark H. Dizer

Design/Layout: *Frances F.M.B. Cardona*

Documenters: *Noelyn H. Curutan*
Rinnah Keziah O. Ramirez
Jerome R. Calicdan

These proceedings have been published by the CEM 46th Anniversary Conference Committee, Center for Educational Measurement, Inc.

33rd Floor Cityland Pasong Tamo Tower 2210 Chino Roces Avenue, Makati City, 1231 PHILIPPINES

Tel: +63 2 8813 3694 to 95
Fax: +63 2 8894 1392
+63 2 8894 5536
Email: inquiry@cem-inc.org.ph

Center for Educational Measurement, Inc.

THE TEST OF EXCELLENCE

FORWARD
Focused Goals. Maximum Results.

46th ANNIVERSARY

VIRTUAL CONFERENCE

PROCEEDINGS

Building a Balanced Assessment System

BEYOND THE NUMBERS

30 SEP 2024 8:00 AM - 12:00 NN VIA zoom

www.cem-inc.org.ph

TABLE OF CONTENTS

PROGRAM

CEM 46th Anniversary Virtual Conference

Building a Balanced Assessment System: Beyond the Numbers

September 30, 2024 | 8:00 AM – 12:00 NN | CEM CPD Online / Zoom

TIME	ACTIVITY
8:00 AM	**OPENING CEREMONIES**
	Welcome Message **Dr. Luis M. Sorolla, Jr., CSEE** *Chair, CEM Board of Trustees Chair*
	Opening Remarks **Dr. Grace H. Aguiling-Dalisay, RPsy** *President & CEO, CEM*
8:30 AM	**CONFERENCE PROPER**
	Keynote Address: Empowering Educators and Stakeholders in Building a Balanced Assessment System **Dr. Jimmy De La Torre** *Professor,* The University of Hong Kong
	OPEN FORUM
	Topic #1: Literacy Forward: Use of CEM's Assessment System in Improving the Quality of Education **Ms. Kathryn M. Tan, MAPsych** *Programs and Development Director, CEM* **Ms. Janet T. Evasco, MA, RPsy** *Operations Director, CEM*
	Topic #2: Innovative Assessment: Exploring Opportunities and Challenges **Dr. Jasper Vincent Q. Alontaga** *Associate Professor, De La Salle University - Manila*
	Topic #3: Implementing Authentic Assessment to Foster 21st Century Learning **Dr. Marie Therese A.P. Bustos** *Director, Assessment, Curriculum and Technology Research Centre (ACTRC)*
	OPEN FORUM
	Moderator Synthesis **Mr. Louie P. Cagasan, Jr., MAPsych** *Assistant Professor, UP College of Education*
11:40 AM	**CLOSING CEREMONIES**
	Closing Remarks **Ms. Iris Lark H. Dizer, MSEdMEv** *Overall Chair, CEM 46th Anniversary Virtual Conference*

WELCOME MESSAGE

Dr. Luis M. Sorolla, Jr., CSEE
CEM Board Chair

Our dynamic CEM President, Dr Grace Aguiling-Dalisay; Our esteemed speakers for this event, my dear members of the CEM Family, participants, friends, ladies, and gentlemen.

It is my distinct honor to welcome everyone to our 46th Anniversary Virtual Conference with our theme, Building a balanced Assessment System: Beyond the numbers. Our theme aptly describe what we are trying to do here in CEM as we endeavor to put meaning in the facts and numbers derived from the various tests and measuring instruments.

As we celebrate the CEM's 46th anniversary, let us glance a while to what we have done so far and focus on what we need to do and accomplish in the years ahead. May everyone think and work together as a formidable team to push forward our goal of improving the educational landscape in this country. We hope to see the day when the Philippines lead other countries in the world in the educational hub, and CEM is one of the strong pillars to achieve this end.

May this activity today be another springboard for CEM towards development, quality and excellence in the educational arena.

On that note, I warmly welcome everyone to this conference.

Good morning!

OPENING REMARKS

Dr. Grace H. Aguiling-Dalisay, RPsy
CEM President & CEO

Thank you, Dr. Luis Sorolla, our Board Chair, for the warm welcome. A pleasant anniversary morning to you and our other esteemed Board of Trustees, Corporators, Conference Speakers, member schools, partners and all participants of this Virtual Conference in celebration of CEM's 46th Founding Anniversary. Your presence is highly appreciated as we mark the end of our Anniversary month.

On September 1, 1978, the Guidance and Testing Division of the Fund for Assistance to Private Education (FAPE) was launched as an independent non-profit educational testing agency called the Center for Educational Measurement (CEM). This was in response to the clamor of private and government schools for the use of the test batteries developed by the FAPE GTD and for technical consultations on test utilization and development.

Since then, CEM has responded to the educational assessment needs of schools and organizations, mindful of the changes in the national and global landscape. In keeping with technological, methodological and conceptual advancements, we have continued to develop assessment tools and intervention programs to support the optimal learning of students, professionals and life-long learners. We have offered annual anniversary conferences on timely topics of consequence such as leading change with data, classroom research for teacher improvement, data-driven plans for Principals, equal opportunities in education through scholarships, measurement for selective recruitment and retention, and educational assessment as a tool for quality education.

The results of the standing of our basic education students in international largescale assessments and national achievement tests are undeniably dismal. However, we note well and with great hope, that the Second Congressional Commission on Education (EDCOM II) has called attention to assessment as a priority agenda in educational reform, and with full support of the Department of Education and the national government. With this considered milestone, CEM forges ahead with advancing the value of assessment towards quality in education for all. We do so in grateful collaboration with our member schools, assessment consortia and stakeholders in the field of quality Philippine education. Our anniversary conference this year on "Building a Balanced Assessment System: Beyond the

Numbers" is one contribution to the attainment of this mission.

As I end my greetings, and before we move to the Conference Proper, allow me to share with you an AVP showcasing our performance highlights for Fiscal Year 2023-2024 covering the months of June 2023 to May 2024. We share these triumphs with you, our partners in education. With your support, we will continue to serve with excellence to ensure our collective success. Let us all watch this.

KEYNOTE ADDRESS:

Empowering Educators and Stakeholders in Building a Balanced Assessment System

Dr. Jimmy De La Torre
Professor, The University of Hong Kong

Professor Jimmy De La Torre, an esteemed academic in the field of educational and psychological measurement, delivered an enlightening presentation on the importance of balanced assessment systems entitled "Empowering Educators and Stakeholders in Building a Balanced Assessment System". Currently a Professor of Human Communication, Learning, and Development at the University of Hong Kong's Faculty of Education, he is widely recognized for his expertise in item response theory, cognitive diagnosis modeling, and assessment integration in instruction and learning. With over 100 articles and book chapters published, as well as more than 40 workshops conducted on cognitive diagnosis modeling across 15 countries and four continents, his contributions to the field are significant. His work has earned him prestigious accolades, including the 2009 Presidential Early Career Award for Scientists and Engineers from the White House, the Jason Millman Promising Measurement Scholar Award, and the 2017 Bradley Hansen Award for Contributions to Educational Measurement. Notably, Professor De La Torre has a personal connection to the Center for Educational Measurement (CEM), having used its products as a student and guidance counselor and having been mentored by Dr. Grace Aguiling-Dalisay, one of his college professors.

Professor De La Torre began his presentation by underscoring the essential role of measurement in fostering improvement, referencing Sir William Thompson's famous quote: "If you cannot measure it, you cannot improve it." He emphasized the necessity of accurate and actionable assessments in informing changes that lead to better educational outcomes. This laid the foundation for his discussion on the concept of a balanced assessment system, which he described as an aspirational framework designed to align various components of education, including curriculum, learning theory, instruction, assessment tools, professional development, psychometric frameworks, and implementation strategies. While no system perfectly embodies all these elements, striving toward balance ensures better integration and outcomes.

A balanced assessment system, he explained, should possess five key features. First, it must be **comprehensive**, offering students multiple avenues to demonstrate their understanding and utilizing diverse evidence sources. Second, it should be **continuous**, with frequent assessments enabling educators to track growth both within and across academic years. Third, it needs to be **efficient**, eliminating redundancy by ensuring each assessment serves a distinct purpose. Fourth, it must be **coherent**, maintaining vertical alignment across different educational levels (national, district, and classroom) and horizontal alignment between curriculum, instruction, and assessments. Finally, it must be **useful**, delivering timely and actionable information that supports informed decision-making by educators, administrators, and other stakeholders. Modern balanced systems also leverage recent psychometric methods and technological advancements to enhance their utility.

Designing and implementing a balanced assessment system, Professor De La Torre noted, is an iterative process that requires several steps. It begins with defining a clear learning framework, developed collaboratively with educators and experts, to establish the underlying theory of learning. Next, existing assessment systems are reviewed to identify effective practices, eliminate redundancies, and address gaps or unmet needs. Based on this review, a comprehensive plan is crafted, aligning assessment activities with the educational goals and learning theories of the institution. This plan should articulate the purposes, types, timelines, and expected outcomes of each assessment, ensuring consistency and coherence. Assessment tools must then be carefully selected or developed to align with the comprehensive plan, with ongoing validation processes to ensure their reliability, fairness, and applicability. Integration with instructional activities is essential, as assessments should support rather than detract from teaching efforts. Furthermore, professional development must be provided to equip educators with the skills to use assessment results effectively. Administrative support is also crucial, encompassing policies and a culture that encourage the use of balanced assessment systems.

To illustrate these principles, Professor De La Torre highlighted NWEA's adaptive online assessment for early literacy as a model example. This system assesses foundational reading skills and oral reading fluency through adaptive technology that adjusts in real time to student performance. Key features include automated scoring of spoken responses using Automated Speech Recognition (ASR) software, alignment with the Simple View of Reading framework and Common Core State Standards, and continuous assessment three times a year (fall, winter, spring). The system provides immediate scoring and reporting, offering valuable benchmarks, raw scores, and instructional recommendations. Additionally, it employs Item Response Theory (IRT) to ensure score comparability across grades and demographics. The system's development involved content specialists and external experts to maintain consistency in cognitive complexity, item format, and content scope, while its technological integration allows educators to access audio recordings for further evaluation. Such systems demonstrate how balanced assessment frameworks can support instructional practices through innovation and precision.

In discussing psychometric frameworks, Professor De La Torre compared three primary models: Classical Test Theory (CTT), Item Response Theory (IRT), and Cognitive Diagnosis Models (CDM). CTT operates at the test level, while IRT focuses on the item level and is commonly used in summative assessments. CDM, in contrast, is a multidimensional model that assesses specific skills, making it especially suited for formative assessments. He noted that IRT assumes a unidimensional latent trait, whereas CDM accommodates a multidimensional latent attribute vector, offering richer diagnostic insights.

Despite their potential, implementing balanced assessment systems faces several challenges. Coordination among diverse stakeholders with varying policies is often complicated by frequent policy changes and communication gaps. Misalignment between assessments, curriculum, and instruction further hinders effectiveness. Limited expertise in modern psychometric methods poses another significant barrier, emphasizing the need for collaboration with external experts. Technological integration also presents challenges, such as ensuring compatibility across platforms, providing adequate access to devices and the internet, and adapting assessments to measure skills beyond traditional multiple-choice formats. Finally, a lack of teacher training on using assessment systems and interpreting results highlights the importance of professional development.

Professor De La Torre concluded his presentation by reiterating the relationship between assessment and learning. He emphasized that assessment alone does not guarantee improved learning outcomes; it must be paired with effective remediation, instruction, and the provision of necessary resources. Comprehensive systems that integrate assessment and learning components are essential for fostering growth. By leveraging modern psychometric methods and technological advancements, educators and policymakers can design balanced assessment systems that align with educational goals, support instructional practices, and ultimately enhance student outcomes. While challenges persist, Professor De La Torre's vision for balanced assessment systems serves as a roadmap for achieving meaningful and lasting improvements in education.

KEYNOTE ADDRESS OPEN FORUM

*After the presentation of **Dr. Jimmy De La Torre**, **Mr. Louie P. Cagasan**, the moderator opened the virtual floor for questions from the participants. Participants were encouraged to ask questions or share thoughts and insights thru the Zoom chat box.*

To ensure accuracy and transparency, the questions and answers from the open forum are presented verbatim.

Question #1

How can we ensure that the different types of assessments are effectively integrated to assist student learning and educational decision-making?

Answer

This is a good question because we are really interested in not only having different types of assessments but making sure that they are actually integrated. This requires a conscious and sustained effort. First of all, we need to take stock of where we are and where we want to go.

This is what we call a comprehensive plan. At least on your own, you need to have a clear understanding of what you hope the assessment system can actually accomplish. However, integration per se is not possible, or it is not the answer if we do not have the requisite assessments in place.

If we want to be able to say we are integrating different assessments, we need to make sure that the assessments are already in place. But for you to have the proper assessments, you have to have the overall goal that allows you to say, what role does this particular assessment play? I think we have to take a step back first, not just say, okay, let me integrate whatever I have, because it may not be possible to begin with, because they are not designed to perform that role. You need to take a step back, look at the assessments, look at the goals, and then fill in the missing assessments.

Note that I only focus on large-scale assessments or standardized assessments. However, assessment, as I described earlier, in the most generic terms, is simply the process of collecting information or data to support your needs. So, this could also happen at the classroom level, where teachers are involved in the development of the assessment; however, teachers need to be mindful that they are not conducting the assessment in isolation. They need to understand how the assessments they are conducting at the classroom level are actually related to the overall goal of the assessment system.

Question #2

What professional development and support structures are necessary to equip educators with the skills and knowledge to implement and sustain a balanced assessment system?

Answer

As you mentioned, and as I also touched on in my presentation, it is actually important to have professional development until such a time that teacher training would take care of this. However, it is probably

impossible in the way we are currently training teachers, and this is actually true not only in the Philippines but also in other parts of the world due to the rapid development of a lot of new technologies.

And as a result, teachers are not necessarily aware or trained to use these technologies and incorporate them as part of their classroom instruction. So, we need to introduce teachers to these technologies, and at the same time, provide them the necessary support they need.

Currently, teachers are already busy. As a result, they are not able to spend more time on other things. Just giving them additional work to do will not necessarily encourage them to adapt a new way of doing things—we need to make their life easier.

I have an example from colleagues in the Netherlands. Teachers are excited to use the system that was introduced to them because it actually helps them with homework assignments. In particular, teachers no longer have to grade the assignments because that can be done automatically online. They welcome the use of technology because they do not have to grade the assignments themselves.

We need to recognize that teachers have needs, as in, the gap in knowledge, and they are already extremely busy and pressed for time. Helping them to address these needs, and supporting them in being able to adopt this new knowledge is a task that is incumbent upon people interested in making sure that the assessments are actually being utilized in a way that is consistent with the overall goal of the assessment system.

Question #3

Which learning theory is most commonly associated with interim assessments, and how can interim assessments be effectively implemented based on this theory?

Answer

The learning theory will depend on the specific domain that you are trying to measure. This should be part of the training, as, teachers need to be informed of the learning theory that will be used for the overall assessment.

When you are doing a summative assessment, it would require touching on the bigger parts of the theory. In contrast, when you are actually interested in interim types of assessments, the minute parts or the more specific aspects of the theory are more useful. Therefore, this type of information needs to be provided to teachers and educators so they can have a clear idea of what the overarching theory is, as well as an understanding of the different components that make up the theory. These components are more relevant to what is happening at the classroom level.

Question #4

Are there local experts on IRT or Rasch model who can help schools and train teachers?

Answer

The answer is yes and no. As Louie mentioned, it has been a long time since I taught IRT in the Philippines, but I know that there are currently other people also who are teaching IRT here in the Philippines. However, fully implementing IRT would require a lot of expertise behind the scenes that would take a long time to acquire. Therefore, we only need the

schools, educators, and teachers to have a high level of understanding of how IRT works. A nitty-gritty understanding does not need to happen at the school level; rather, this is needed at the level where large-scale assessments are being developed.

So, yes, there are experts in the Philippines, although still very scarce, but this does not have to be within each school. At the school level, a clear understanding of why the scores are comparable and how the different score types--interim, summative, or formative--can be used would suffice.

At the school level, it would be enough for teachers and administrators to understand the proper use of the reported scores without going to the nitty-gritty of item response theory.

Question #5

What can you advise educators who use grade results as the sole source of data in terms of interventions? Is it valid? What is your take on this?

Answer

This is something worth highlighting. The fact is you should never use a single score to make important decisions for various reasons. Many times, test that do not necessarily align with what is happening in the classroom is used. Thus, decisions based on such a score may be very misleading. However, even when tests are actually designed to align with classroom instruction, they can only provide a snapshot and a subset of what the students know and can do.

This is why teachers are encouraged teachers to supplement test scores with additional information--in addition to test scores, teachers have other sources of information about their students that they can use to make important decisions. The challenge is to put together the different pieces of information to arrive at a more comprehensive picture of what the students can and cannot do. From there, we decide what the appropriate next steps should be.

Question #6

What measures can we take to ensure the security and integrity of our assessment data? And how can we balance the need for standardized testing with the need for more personalized, student-centered assessments?

Answer

It depends on what you mean by data security. I can answer this question in different ways. For one, data security could mean how trustworthy the test will be in the future, and this is more or less relevant when you are talking about high-stakes tests. If the decisions involved are quite high stakes, you need to be more careful about the security of, not just the data, but also the assessment itself.

I am actually not sure what we mean when we talk about data security. Do you mean data privacy? If so, I think this is something that needs to be done at the school or at the district level because there are typically different policies involved regarding what scores can be shared with different individuals. Hence, you need to understand the policies or the limitations of what you can or cannot share.

As I mentioned earlier, a single test cannot cater to all the needs of different stakeholders. If you have a summative assessment, which gives you a broad stroke

of what students can do over a longer period of time, it will not be as specific as a test that is given more frequently.

To this end, it is possible to have some type of standardized (let me put standardized in quotations) assessment that happens more frequently. This is what I am advocating when it comes to formative assessment. You need to deliver them more frequently and report the scores in a timely manner so that the results can be integrated into the student learning process.

In addition, teachers need to be also trained to develop classroom assessments that fit into the bigger picture. Alternatively, this is a resource that can be provided to them. For example, instead of developing their own assessment each time, they could be given access to an item bank for the purpose of conducting formative or interim assessments.

As I have said, these are activities that are not necessarily independent of each other. These activities are, in fact, designed to complement each other. There is certain information that large-scale assessments, standardized tests, or summative assessments cannot provide.

This is where teachers can step in, but they need to be provided with the necessary resources. Teachers interested in finding out more about their students should not be left to do everything on their own. This would be a very demanding task for teachers because, as we said earlier, they are already busy. Directing or giving them access to resources can actually make the adoption of certain practices more feasible in the future.

Question #7

What can you say about teacher-made diagnostic tests?

Answer

I think they play an important role in the entire balance assessment system. Teachers, who give diagnostic tests, are on the ground, so they know what is actually happening.

But at the same time, as I mentioned earlier, they need to be oriented to the bigger picture. They need to understand that, when they create their own diagnostic assessments, these assessments have to fit into the bigger picture. This can prove too demanding for teachers because diagnostic assessment does not happen just once a year.

It needs to happen when you move from one material to another or after one chapter or one quarter. You need to diagnose where the students are to make sure that materials can be adapted, and existing gaps addressed. In other words, you need to be able to come up with empirical evidence that you can use to adapt your instruction.

To this end, teachers have an important role to play, but this is a very demanding task on top of the responsibilities that they already have. Therefore, any resources that teachers are provided with can help simplify their work.

Question #8

How can formative assessments be effectively integrated into a balanced assessment system?

Answer

As I have said, it starts with recognizing that every assessment plays a role in the bigger picture. Recognizing that formative assessment plays an important part in a bigger system is a prerequisite to creating a coherent system.

Using diagnostic or formative assessment is like going to a doctor. A doctor might tell you what potentially could be a problem that you need to address; however, you do not stop with the assessment. There should be a corresponding remedial step that one should take to address the specific issues at hand.

To a large extent, this is not something that all teachers can readily do. This is why teachers need help by working with other teachers, seeking the help of experts, and making use of available resources. If this cannot happen at the school level, teachers can seek help at the school district, regional, or national level, where, hopefully, greater expertise and more resources are available.

However, the higher up you go, the more removed it is from the actual classroom practices; thus, the available help or resources may be one-size-fits-all solutions that do not really work across all situations.

For example, adaptive testing represents a reasonable and efficient solution for formative assessment purposes. However, we still have many schools in the Philippines that do not have access to electricity, computer devices, or the Internet. For these schools, adaptive testing is not a viable solution. We need to recognize that, although effective in some contexts, some solutions may not be applicable in other contexts; hence, different, more suitable solutions need to be devised.

TOPIC #1:
Literacy Forward: Use of CEM's Assessment System in Improving the Quality of Education

Ms. Kathryn M. Tan, MAPsych
Programs and Development Director, CEM

Ms. Kathryn Tan serves as the Director of Programs and Development at CEM. Throughout her career at CEM, she has been committed to leveraging assessment as a tool for enhancing educational outcomes. Her work encompasses analyzing CEM test data to extract valuable insights related to teaching and learning, as well as assisting schools, individual researchers, and various organizations in utilizing CEM test results for research and institutional planning purposes.

Ms. Tan delivered a presentation titled "Assessment as a Lever for Quality Education," which was structured into two main components. The first part focused on elucidating the role of assessment in enhancing the quality of education, while the second part outlined how CEM seeks to implement this concept.

In the first part of her presentation, Ms. Tan defined quality within the context of education, exploring various interpretations and perspectives on what constitutes quality in educational settings. She emphasized that the concept of quality education can be understood as judgments about the inputs, processes, outputs, and outcomes of education. Additionally, she defined assessment as a systematic process utilized to measure or evaluate characteristics or performance, and mentioned that standardized testing is designed to implement consistent procedures for content determination, administration, scoring, and reporting of results.

Ms. Tan further elaborated on the pivotal role of assessment in education, highlighting its importance in supporting decision-making across diverse educational contexts. She identified several specific contexts for this decision-making. Beginning with teaching and learning, she noted that formative assessment is the appraisal of student performance primarily to enhance student learning, by improving ongoing instruction and guiding student effort. On the other hand, summative assessment is the 'final' appraisal of student performance to certify what has been learned but it can also support learning, by stimulating consolidation of

knowledge and rehearsal of strategies essential for the development of expertise.

In the context of guidance and counseling, the use of assessments revolves around helping students make decisions about educational, career, or life plans. She further noted that in the context of student admissions, the purpose of testing is to maximize the selection of students who have a high likelihood of performing well in a particular program of study. For program and planning, she mentioned that assessments can also be used to describe and appraise programs as well as people. Lastly, she emphasized that theoretical and applied research and education also make use of assessments to study various educational processes, outcomes, or contextual variables.

In discussing the relationship between assessment and quality education, Ms. Tan highlighted scenarios in which decision-making pathways utilize assessment as a catalyst for enhancing educational outcomes. Specifically, she noted that assessment plays a critical role in informing planning and decisions related to improving the inputs of education, as evident in selection and resource allocation processes. Furthermore, assessments enhance educational processes by providing feedback that supports both teaching and learning, as well as guiding learners in their educational and career choices. Additionally, assessments serve to clarify the outputs and outcomes of education by documenting the achievement of educational objectives and providing a means to demonstrate qualifications. This approach can significantly broaden access to educational opportunities for individuals who might otherwise remain unrecognized.

Ms. Tan also emphasized that in the context of decision-making, standardized assessments offer significant advantages. Notably, these assessments undergo rigorous review and pilot testing, which are instrumental in determining the appropriateness of the content and difficulty level for students across various schools.

In the second part of her presentation, Ms. Tan shared on how CEM is committed to enhancing decision-making processes in various educational contexts through the implementation of meticulously developed and standardized assessments. She emphasized the transformative impact CEM aim to achieve through these services. She provided an overview of the assessment solutions offered by CEM, which include a diverse range of tests designed to evaluate competencies in basic education, as well as assessments focused on career decision-making. Additionally, she highlighted several assessments for admission to advanced academic programs, including the Philippine Aptitude Test for Teachers and the Panukat ng Pagkataong Pilipino developed by Dr. Carlota.

In terms of teaching and learning, she noted that CEM achievement test reports provide information that directs teachers' attention to student competencies and learning gaps or difficulties and helps schools use the results from these assessments through the conduct of test interpretation and data utilization seminar workshops where administrators, teachers, and guidance counselors learn how to interpret our test scores and use these to identify targets and plan in curriculum program or instructional interventions. Through these services, she shared that CEM hopes to create in schools the practice

of continuous assessment using both teacher-made tests and external assessments like the CEM tests, examination of these results, and action based on assessment findings, and subsequently, may to contribute to improving student learning.

For guidance and counseling, she highlighted that CEM provides information on applicants' career aptitudes and interests, plus an online database of occupations to serve as tool for exploring career options during guidance counseling. She shared that the seminar workshops are also provided to help guidance counselors interpret and use the test data to give counseling advice and create career guidance programs. She underscored that CEM seeks to support educational and career decision making aligned with interests and aptitudes through these tools and activities. And in the process, CEM hopes to contribute to enhancing applicant program fit for students, alignment of their educational paths with their interests and aptitude, and satisfaction and successful performance in their chosen paths.

Ms. Tan also shared that for admissions and scholarships, CEM tests provide information on aptitudes or relevant competencies. According to her, the information is intended to help make the merit basis of the selection process rational, ensuring that all applicants, regardless of background, have an equal chance to demonstrate the necessary aptitudes or competencies. And with this, CEM aims to contribute to enhancing applicant program fit or readiness of candidates for the next level of studies.

Ms. Tan concluded her presentation by highlighting that assessment informs planning and decisions related to improving inputs, processes, outputs, and outcomes in education; and that it is not merely an auditing mechanism but can positively influence development when well designed.

With emphasis she mentioned that ultimately, through these decision-making pathways, assessment serves as a lever for quality education, enhancing student outcomes and supporting learners effectively; and given key role that assessments play in decision making, CEM takes great care in ensuring that the information we provide to schools can be relied on.

Ms. Janet T. Evasco, MA, RPsy

Operations Director, CEM

Ms. Janet T. Evasco is a seasoned expert in educational measurement and psychology, is the Director for Operations at the Center for Educational Measurement (CEM). She holds a Master of Arts in Psychology from the University of the Philippines Diliman. With 18 years of experience, she was instrumental in creating and overseeing major national assessments like the National Medical Admission Test and the Philippine Law School Admission Test, ensuring their alignment with high standards. Ms. Evasco's influence extends beyond technical roles to strategic leadership and active engagement with professional bodies for continued professional development. Her discussion "Ensuring Validity and Integrity in Educational Testing, A Comprehensive Approach to Ethical and Reliable Testing Practices" focused on what CEM does to ensure credible results are provided to clients.

Ms. Evasco started with the importance of validity, reliability, and integrity in testing. According to her, validity is essential in making sure that test scores accurately reflect their intended purpose. She added that valid test scores give a true picture of the abilities or readiness of the test taker. Without validity, assessment results would be misleading, contributing to poor decision-making in academic or professional settings. She mentioned that valid measurements are generally reliable which maintains the test's high credibility. Thus, it is necessary to maintain the integrity of the test to obtain scores under secure and fair conditions. In addition, standardization in procedures helps in achieving accuracy and comparability in the interpretation of scores.

Ms. Evasco highlighted that the CEM assessment products undergo standardized procedures for test development, testing directions and conditions, and scoring, regardless of testing modality, including pen and paper and online computer-based tests. She explained the detailed process of ensuring test integrity followed by CEM: (1) Test Development Process: CEM has a team of licensed psychometricians and subject experts who follow a systemic approach in developing, evaluating, and trialing test items at different stages to meet the high-quality standards required in the testing industry, (2) Test Administration: This process includes clear and precise directions given by examiners who carefully follow a comprehensive Test Administration Manual (TAM). The TAM is a set of policies and procedures on standardized test administration, securing copyrighted test materials, and providing verbatim instructions before, during, and after testing, (3) Training and Support of Testing Staff: Test administration experts are led by registered psychometricians, who also train the testing staff to develop their proficiency in test administration and

adherence to standards. Through online platforms, constant coordination is also ensured among testing personnel, (4) Creating a Comfortable Testing Environment for Examinees: CEM arranges a comfortable and supportive testing environment for examinees so that they may perform at their best. Examinee orientation for online computer-based tests is held to discuss prohibited materials or behaviors and hardware requirements. A demo test is provided to familiarize candidates with the platform. For paper-based tests, CEM favorably responds to schools' practice requests in shading answer sheets to help examinees gain confidence. On either test delivery method, human proctors monitor test sessions and assist with test takers' concerns without causing undue advantage to anyone. Clear instructions and sample test items are also given. Some accommodations may be available for some test takers for a stress-free experience, (5) Preventing Fraudulent Means of Obtaining Scores: CEM ensures that test results are accurate by preventing and detecting fraudulent methods and implementing random sampling designed for the administration of the items for online exams. Continuous monitoring minimizes security breaches and trained proctors handle testing irregularities, and (6) Scoring, Data Security, and Confidentiality: Wherein, a dedicated team trained in machine-aided scoring and quality control adheres to established protocols. Reports on individual and group performance are provided with a Score Interpretation Guide (SIG). The SIG describes what the test covers, score meanings, and interpretation guidelines. It also details the reliability of scores and their intended uses. Achievement tests report category-specific performance by grade level. The criterion and norm-referenced scores assist in decision-making and understanding student profiles. Test interpretation seminars are offered for deeper understanding and maximizing data use. CEM also ensures data privacy by maintaining confidentiality and executing data-sharing agreements with partners. Data collection, handling, storage, and disposition are enhanced by a designated data protection officer and IT team.

To conclude, Ms. Evasco addressed CEM's commitment to standards. Its dedication to upholding educational and psychological testing standards to ensure test quality and integrity continuously aligns practices with international standards. CEM operates under the International Standards for Educational Assessment Organization (IAEA) which CEM is a proud member, and the Joint Standards for Educational and Psychological Testing (2014), by the American Educational Research Association (AERA), American Psychological Association (APA), and the National Council on Measurement in Education (NCME).

TOPIC #2:
Innovative Assessment: Exploring Opportunities and Challenges

Dr. Jasper Vincent Q. Alontaga
Associate Professor, De La Salle University - Manila

Dr. Jasper Vincent Q. Alontaga is an Associate Professor under the Department of Educational Leadership and Management, at De La Salle University – Manila, handling educational technology and other professional education courses for over 20 years using blended learning delivery. He specializes and serves as a resource speaker and consultant on technology integration, digital leadership, flexible learning and online course development. His presentation, "Innovative Assessment: Exploring Opportunities and Challenges," provided a comprehensive exploration of how technology can enhance assessment practices in education, making them more relevant, effective, inclusive, and fair for all learners.

Dr. Alontaga began his presentation by sharing his insights about Innovative assessment. He explained that assessments for 21st century students must be innovative, emphasizing the integral role of technology in the assessment process. However, he also cautioned that while technology offers opportunities for innovation, it can become a source of distraction or even "destructive innovation" if not deployed properly and effectively. He urged the audience to explore the opportunities and challenges of technology in assessment and reflect on how these tools can be integrated within individual contexts. Dr. Alontaga highlighted the vast differences in technological infrastructure and resources across schools, grade levels, and regions, even within the same city or across the Philippines.

He outlined three key elements essential for effective assessment: the students who take the assessments, the teachers who design them, and the systems or technology used for their design and delivery. Dr. Alontaga focused on the third element—the systems or technology—prompting the audience to ponder an important question: "How can we leverage systems to improve assessments, including their security and integrity?" He emphasized that any innovation in assessment should enhance the overall process, enabling students to better

demonstrate their knowledge and performance. Additionally, he stressed the importance of distinguishing between *security* and *integrity* in assessments. While technology is a natural step for institutions aiming to adopt innovative practices, it is critical to ensure that these advancements not only protect the assessment process but also uphold its integrity.

Dr Alontaga introduced the SAMR model, developed by Ruben R. Puentedura, as a framework for integrating technology into educational practices. The model outlines four levels of technology integration: Substitution, Augmentation, Modification, and Redefinition, representing a progression from basic usage to transformative educational experiences.

He explained that everyone begins with no technology, and moving forward requires courage to embrace its role in education. The journey typically starts at the *Substitution* level, where technology is used as a direct replacement for traditional tools, allowing educators to explore and familiarize themselves with new methods. As confidence grows and educators delve deeper into technology, they progress to the *Augmentation* stage. At this level, technology continues to act as a substitute but now includes functional improvements that enhance the learning process. Dr. Alontaga highlighted that the final two stages, *Modification* and *Redefinition,* represent the transformational phases of technology integration. These stages involve reshaping and reimagining educational practices, enabling activities and assessments that were previously inconceivable.

He emphasized the importance of a gradual progression through the levels of the SAMR model. He advised against jumping from a no-technology approach straight to the *Redefinition* level, explaining that educators should first gain experience with *Substitution*, *Augmentation*, and *Modification* before attempting full transformation. Skipping these stages, he explained, can lead to two possible outcomes: either teachers will manage to navigate and adapt to the complexities of *Redefinition* with significant effort, or they may struggle and feel overwhelmed as it could result in viewing technology as a distraction rather than an innovation in assessments. To avoid this, he encouraged teachers to build confidence step by step, exploring practical examples and addressing challenges and opportunities within each level of the model. This approach ensures a more effective and meaningful integration of technology into assessments.

Dr. Alontaga shared various practices utilized in their system, such as the use of Canvas LMS. He highlighted the platform's capability to provide item analysis reports, including the discrimination index and response breakdowns, which assist educators in revising, retaining, or replacing test items to enhance assessment quality. He also discussed educational tools like Kahoot, FlipQuiz, Wizer, and Edpuzzle, which integrate multimedia and interactive formats to make assessments more engaging. He emphasized that motivational assessments reduce concerns about assessment security, as engaged students are less likely to resort to dishonest practices.

Moving on to the Substitution and Augmentation levels, Dr. Alontaga emphasized the importance of careful implementation. He recommended

practices such as digital honor pledges, passcodes, and time limits for online exams. However, he cautioned against the misuse of test banks, stressing that randomized items must align with the test's table of specifications to avoid unnecessary difficulty. He also pointed out the utility of tools like AI-generated exam items, LockDown Browser, and online proctoring tools to enhance assessment integrity but advised educators to remain vigilant about their limitations.

Transitioning to the Modification and Redefinition levels, Dr. Alontaga encouraged educators to reflect on how AI can support tasks and assessments. He urged caution in interpreting plagiarism and AI reports, emphasizing that these tools are not definitive in determining the authenticity of outputs. He also advocated for empowering students by involving them in self- and peer-assessment, which helps them understand the characteristics of good performance, fosters accountability, and facilitates group feedback on roles and contributions.

In his conclusion, Dr. Alontaga shared best practices for implementing innovative assessments. He advised educators to be patient when trying new types of assessments, to remain flexible in accommodating students with technical or other challenges, and to clearly communicate assessment parameters to ensure alignment with student expectations. He also emphasized the value of seeking assessment testimonials to inform the redesign and improvement of future assessments.

TOPIC #3:
Implementing Authentic Assessment to Foster 21st Century Learning

Dr. Marie Therese A.P. Bustos

Director, Assessment, Curriculum and Technology Research Centre (ACTRC)

Dr. Marie Therese Angeline P. Bustos is a Professor of Special Education at the University of the Philippines College of Education, specializing in special and inclusive education research. Her work includes research reviewing the K-12 curriculum, incorporating disability in the conditional cash transfer program, baseline disability studies in Indigenous and Muslim communities, and assessing DepEd schools' readiness to provide services for children with disabilities in marginalized communities. She was previously the Dean of the UP College of Education and currently serves as the Philippine Director of the Assessment Curriculum Technology Research Center. Dr. Bustos chairs the Technical Committee on Special Education and previously served on the Technical Panel on Teacher Education at the Commission of Higher Education.

In her presentation "Implementing Authentic Assessment to Foster 21st Century Learning," she discussed how authentic assessment can allow learners to apply their learning to real-world problems and develop 21st century skills such as critical thinking, communication, collaboration, and technology literacy.

Measuring Learning Outcomes

Dr. Bustos began by presenting the results of the Assessment, Curriculum, and Technology Research Center (ACTRC) review project focused on Senior High School. The central question of the project was, "How should student learning outcomes be measured in Senior High School?" Consultations with various stakeholders revealed differing perspectives. Some advocated for standardized tests to determine if students achieved the targeted outcomes, while others suggested that TVET national certificates could serve as evidence of meeting these goals and would have already shown that they have met the targeted outcomes. Others supported performance-based assessments. There were also calls for a combination of written tests and performance measures. Dr. Bustos emphasized that evaluating the success of Senior High School should also involve analyzing graduates' performance across various exit pathways.

Dr. Bustos raised that the issue on how to assess outcomes also exists in teacher education, raising the question: "What would be a good measure of learning outcomes for teachers?" She questioned whether passing the Basic Licensure Examination for Teachers (BLEPT) alone suffices and whether holding a license adequately reflects teacher preparedness. She referenced the Philippine Business for Education (PBEd), which has pushed for a review of the licensure exams due to disappointing outcomes from teacher education institutions.

This particular problem is not just true for the Philippines, but also true in other places such as in California, as cited by Dr. Bustos. The state introduced a teaching performance assessment, which means aside from passing tests, those who want to get credential need to have to submit evidence of competencies. As an example, an evidence of what they know in terms of learning and lesson planning are required. However, CalTPA (California Teaching Performance Assessment Seminar) has been met with negative reactions among teachers for being expensive and tedious.

Dr. Bustos highlighted the expectations of employers and parents, who adhering to the saying "the proof of the pudding is in the eating," demand assurance that teachers are well-equipped with the skills and knowledge necessary to guide students effectively. This expectation highlights the importance of looking beyond standardized tests as the sole measure of teacher readiness. She emphasized that authentic assessments are essential in providing additional proof of teachers' capabilities.

Moreover, Dr. Bustos referenced the work of Marion and Buckley on teacher evaluation, stressing institutions who graduate teachers need to be accountable in terms of the kinds of graduates that they churn out. One approach to ensuring teacher quality is to establish student learning objectives that clearly outline how teacher training programs are designed. For instance, a student learning objective should clearly state the goal, strategies that will be used to achieve the goal, performance targets for both students and teachers, and a thorough explanation of the assessments that will be used to evaluate the learning goals. Authentic assessment is then introduced as a crucial component of this evaluation cycle for teachers.

Authentic Assessment

Dr. Bustos cited Mueller's (2018) definition of authentic assessments as performance assessments that use real-world tasks and contexts. It values the student in the teaching and learning process. Such assessments are not only a form of testing but also enhance the teaching-learning process itself by engaging students in learning.

Characteristics of Authentic Assessment

Dr. Bustos referred to Wiggins' (1989) framework for authentic assessment, highlighting three key characteristics: structure and logistics, intellectual design features, and grading and scoring. Authentic assessments are often public, involving an audience or a review panel to evaluate students' work. They avoid unrealistic or arbitrary time constraints and are structured more like portfolios than single, one-time tasks. They also require some collaboration with others, supporting group efforts and collective learning. They are able to provide students feedback in

relation to their achievement of certain goals.

For intellectual design features, there is the need to clearly state what is expected of the students. Authentic assessments focus on essential tasks which are constructed to point the student towards more sophisticated use of the skills and knowledge. They are contextualized and involves complex intellectual challenges. Dr. Bustos referenced Frederiksen (1984), who emphasized the value of "ill-structured tasks" in assessments, which are often excluded from standardized testing but are vital for encouraging real-world problem-solving and collaboration. In authentic assessments, grading centers on criteria that assess the essential skills and knowledge in reference to performance standards. This multifaceted scoring system reduces subjectivity and ensures a high degree of interrater reliability as rubrics are utilized by multiple trained judges who apply agreed-upon criteria.

Authentic Assessment and Project-Based Learning

Authentic assessment shares characteristics with project-based learning, where both approaches start with a challenge and engage students in meaningful tasks, as Dr. Bustos noted.

Comparison of Authentic Assessment to Traditional Assessments

Dr. Bustos discussed that a comparison can be made between traditional assessments and more authentic assessments. The ill structure mentioned is similar to the unpredictability observed in authentic assessments, which is not present in standardized testing because everything is predetermined.

Expectation of a task

When students are expected to tap into higher-order skills, it is essential to allow them to engage in exercises that are more unpredictable and real, rather than contrived. Dr. Bustos cited Cumming and Maxwell's (1999) first order expectation which is that students know the facts, figures, concepts, principles, and important information about the subject. They should have developed knowledge and skills necessary to accomplish the task. The second-order expectation is for students to be able to put all of these together and to be able to apply this to a situation. They are immersed in the activity and the performance of actual behaviors that are relevant to the situation and subject area. This approach encourages students to apply, evaluate, and synthesize their knowledge and skills.

Designing Authentic Assessments

Dr. Bustos outlined the critical steps in designing authentic assessments, emphasizing the importance of identifying the prerequisite skills and necessary skills needed for students to create a product. She emphasized that authentic assessments require sufficient time for development, as they must be well thought out and carefully designed.

According to her, in terms of design, it is essential to define the construct clearly, distinguishing what is relevant from what is irrelevant. It is important to ensure that the definition of the construct is comprehensive, as there is a tendency to focus on just one aspect of an idea or concept rather than examining it in its totality. Generalizability is a key consideration when evaluating authentic assessments. Constructs have to be carefully analyzed and the knowledge and

skills that are required to be able to attain that learning goal have to be specified. Dr. Bustos emphasized that it is not simply thinking of interesting or engaging activities. Furthermore, the tasks and responses must fully represent the different facets of the goal to make sure that not only one aspect of the construct is considered. To ensure that an authentic task is generalizable, it is important to define it clearly, develop a good set of rubrics, and train raters effectively. In addition, she recommended to implement several assessment tasks, to gain a more holistic view of a student's performance.

As cited by Dr. Bustos, Spady's (1994) demonstration of mountain of performance, the goal is to create assessments that do approximate things that happen in real life, because that is the only time when learning becomes transformative. While it may be easier to design structured task performances that clearly distinguish the necessary skills and knowledge, such an approach can feel contrived.

Dr. Bustos mentioned that as we prepare our students for the complication of life, we need to be able to think about things that they will do that could best approximate those challenges, and in so doing, prepare them better. This requires thinking about assessments not as one-time events but in terms of long-term development. Assess with the long-term significant outcomes of the program in mind, not just subject-specific outcomes. Furthermore, assessment should provide a comprehensive picture of the student. According to her, rather than viewing it as a single photo, it should be conceptualized as a photo album. This is why authentic assessments often include portfolios or exhibitions. It is not just one assessment, but several assessments that will enable a better picture of students and what they can do.

She added that when creating authentic assessments, it is crucial to be mindful of the standards set in the curriculum. The tasks designed should directly respond to those standards. Clarity about the criteria is essential. She referenced Mueller's (2018) statement on having a clear criteria before creating rubrics. Dr. Bustos identified that the K-12 Matatag curriculum provides clear guidance to distinguish what students need to know and what they need to be able to do. Moreover, collaboration with colleagues is highly advised in the process of designing authentic tasks. After identifying tasks, the next step is to define what good performance looks like. This involves determining the evidence needed to show that students have acquired the targeted outcomes.

Guidelines for developing rubrics

Dr. Bustos discussed the guidelines for developing rubrics. She stated that rubric tells us how well the students perform. Typically, it includes descriptions of exemplary performance, dismal performance, and the average middle. Different styles can be used for rubrics, but the outcome will be levels of performance and a description of the performance per category. She referenced the experiences of ACTRC and the study of HK PolyU (2018) with the importance of an extensive pilot for rubrics to ensure its continued relevance.

Authentic assessments to foster 21st century learning

Dr. Bustos stated the four categories of 21st century skills according to the Department

of Education: information and media technology skills, learning and innovation skills, communication skills, and life and career skills. She stated that is important to recognize that 21st century learning encompasses more than just these skills. It includes an examination of the core subjects that students are taking, awareness of learning exits, and the workforce skills being targeted. Additionally, the readiness of students to use technology and the ability to collaborate with peers are essential components of 21st century learning. She mentioned that in ACTRC, a study was made focusing on collaborative problem-solving utilizing technology.

To ensure students develop skills such as critical thinking, problem solving, collaboration, technology literacy, creativity, and innovation, Dr. Bustos emphasized the role of authentic assessments. She raised a question if the 21st century skills can be effectively achieved using standardized tests to which she referenced Frederiksen's note on standardized assessments' tendency to exclude less efficient tests. This often results in the removal of ill-structured problems, which means that certain essential skills may not be adequately tested to ensure having good items.

Evaluating 21st Century Skills with Situational Judgment Tests (SJTs)

Dr. Bustos discussed the potential of situational judgment tests (SJTs) as tools for evaluating 21st-century skills. Commonly used in HR practices, these assessments provide a scenario with multiple options, allowing students to select the most appropriate response. However, she emphasized that if this approach is chosen, it is crucial to use critical incidents and have a good understanding of a good scenario. Additionally, the options provided must be thoughtfully designed, avoiding leading choices. Attention to construct validity and the validity of the options are important.

Situational Judgment Tests and SOLO Taxonomy

Dr. Bustos highlighted the application of the Structure of Observed Learning Outcomes (SOLO) taxonomy in assessments. SOLO Taxonomy provides a framework for looking at how different students develop their ideas, ranging from having "no idea" to having "extended ideas" or from basic to complex understanding.

She discussed ACTRC's project, Building Basic Assessment Competence of Teachers, which utilized SJTs wherein items with four response options were designed, but upon analyzing the category response curve, they discovered that some options did not perform well. They failed to accurately represent specific levels within the SOLO taxonomy. The hypothesized categories in SOLO items need to be empirically validated to see if these are functioning. She cited Cagasan and Santos' (2024) caution in using it for grading purposes especially if the SOLO items are not empirically validated.

In conclusion, Dr. Bustos underscored the transformative potential of authentic assessments in fostering meaningful learning and preparing students for real-world complexities. She emphasized the importance of designing assessments that align with curriculum standards, provide comprehensive evaluations, and support the long-term development of both students and educators.

PLENARY SESSIONS OPEN FORUM

The virtual floor was once again opened by the moderator, Mr. Louie Cagasan, for another open forum after the plenary speakers' presentation. Participants were encouraged to type in their questions in the Zoom chat box for the speakers to address. The moderator also welcomed thoughts and ideas for significant learnings from the session.

To ensure accuracy and transparency, the questions and answers from the open forum are presented verbatim.

Question #1

How can we ensure the test is aligned with the learning outcomes and accurately measures the knowledge and skills being assessed?

Answer of Ms. Kathryn Tan

In reference to CEM tests, from the test user side, test specifications matching is practiced. Before subscribing to our achievement tests, schools are encouraged to discuss with the testing coordinator to see how well the test coverage matches the curriculum they are implementing. This ensures that the results will align with the objectives set for the grade level and subject area. In terms of test development, we ensure that the test content is appropriate by first clarifying its purpose and then choosing the appropriate basis for the content. We work with subject area specialists and content experts to prepare the test framework and conduct extensive qualitative reviews of the content before it is even pilot-tested in schools. At this point, we gather statistical information to determine the difficulty and level of discrimination. The test goes through a rigorous process to ensure that it covers the appropriate content.

Question #2

How can we ensure the validity and integrity of educational testing while addressing academic dishonesty?

Answer of Ms. Janet Evasco

We have a dedicated team in test administration to ensure that test materials, especially the items, are not exposed. Our well-trained team, in the case of our OCBT (Online Computer Based Testing), ensures that a test taker can submit their identity. We usually ask for valid identification so that each examinee is given entry to the testing platform as the person they declared to be. We use a secure platform for online computer-based testing with AI capability to detect behaviors potentially indicating cheating. We also have human proctors who verify and make the final decision on whether the data collected from the test taker, flagged for problematic behavior (e.g., using a cell phone), is valid. In this aspect, we secure the testing process and ensure the data we collect has integrity.

Question #3

How can innovative assessments be designed to better measure complex skills such as critical thinking, creativity, and collaboration compared to traditional testing methods?

Answer of Dr. Jasper Alontaga

That would primarily be in the rubrics. It is very critical that when moving towards innovative assessment and the higher forms of innovative assessment projects, you need to be able to target the critical skills you want including creative and critical thinking. Good rubrics lead to better innovative assessment.

Question #4

If we are going to reteach the models and principles of educational assessment to Gen Z pre-service teachers without talking about the classic old principles, what and how would it be?

Answer of Dr. Therese Bustos

What they need to be very clear about is that the difference between formative and summative assessment is the purpose. Sometimes we just look at formats—what is a good format for formative assessment, a good test for summative assessment? An assessment, whatever the form is can either serve as formative or summative. Teachers need to be very clear on that because it is the purpose of the assessment that determines if it is formative or summative, not the format of the test. If you're using it for grading and promotion purposes, and you want to know what students learned over a period of time, then definitely that is summative. And if you want to use the same instrument again, scores from the same instrument to improve the teaching-learning process, then you're using it formatively. We don't only collect data because sometimes teachers are so tired, teaching the entire day, and then have things to check. Many times, they'll create tools to collect data, but they don't have time to interpret data and use data to improve teaching. If there is anything that I would like to change in the way we're teaching assessment to our students, especially the ones in teacher education institutions, there is a need to focus on not just the collection of data, but the interpretation of data and using the data.

Question #5

What should be the focus of the curriculum to address the futures design thinking model?

Answer of Dr. Therese Bustos

When we think about 21st-century learning and 21st-century skills, we're preparing our students for an uncertain future, a future that they don't know. We provide them with skills or teach them skills that will prepare them for that uncertain future. This is now the focus of the curriculum. That means futures design thinking models, while they could be a premise on content areas, have to go beyond the content of the subject matter, truly tapping into those 21st-century skills.

Industry has deplored the ability of our students in terms of problem-solving, communicating, and collaborating. Our graduates, according to them, are not good at these skills. This means, that if we are to prepare a good curriculum, it should not just focus on the content of different subject areas, but must infuse those key skills necessary for joining the workforce.

Question #6

In relation to the implication of Dr. De La Torre's presentation earlier, that assessment should be done in a continuum, and in his example where he pointed out standardized testing being given at the start, in the middle, and at the end of the school year.

Answer of Ms. Kathryn Tan

My biggest takeaway from the keynote presentation is the importance of making assessments coherent and being clear on your educational goals from the start. Looking at assessments as evidence of learning allows you to measure or monitor the achievement of those goals. In that perspective, the standardized tests we offer measure what students know and are able to do by the end of the school year. It provides evidence that can be examined in relation to their school results, in terms of their mastery tests, forming part of the photo album of performances that show what students can do. If administered at the start of the year, the assessment can provide baseline information regarding students' initial knowledge and skills. This can be compared to their end-of-year performance to gauge improvement or changes in learning over the year. For interim assessments, context is also an important consideration in determining whether to develop such tests. One challenge we've encountered is the variation in curriculum content and the timing in which schools cover such content, making it challenging to choose a reasonable basis for comparison. This has been one consideration of why we don't currently have interim assessments, but we continue to explore this as other assessment models become available.

Answer of Dr. Jimmy De La Torre

Sometimes the goal is actually very grand that we cannot really achieve them in one fell swoop, we have to really work on it slowly in a sense that what can be done at this point, but we need to be mindful of the eventual goal. In that respect, as Ms. Kathryn was mentioning about interim assessments, and there are places where technology can play a very important role here. If we can make a system available to the teachers where the teachers can even choose the topic where they could limit the area that they will be tested on, and then the system will create the report for the students. This could be done very quickly, on a weekly and even on a daily basis where the students will go in for five minutes just to get tested, and that information will be summarized and provided to the teachers.

Then there are many ways where technology can play an important role. This is something that could actually be what we can aspire for in the future, where how we can seamlessly integrate technology in the assessment process. In some of them, the stakes would be different, so it may not be as stringent as a high-stakes test, but this is more for everyday use, so this could be less stringent in terms of the psychometric requirements.

At the same time, it serves a different purpose that's still part of the big challenge that we need to do. As I've said, the bigger challenge still is can we help in the learning process? To what extent can we even deploy learning material, instructional material, so that once we give the assessment, we can even recommend or provide them with the resources that they need to actually address their needs.

Question #7

How much time does it take for CEM to release test results so that teachers as well as administrators can use these as baseline data? Assuming that a school gives one of the tests at the beginning of the year.

Answer of Ms. Janet Evasco

Our turnaround time for results delivery ranges from two weeks to four weeks but CEM offers flexibility in terms of delivery of results so that it can be used by our stakeholders. We offer that through email. If they allow us to send it through email, then it can be sent to them through email. In relation to their need for earlier results, it is important that they schedule with us earlier the test administration.

Question #8

If we digitalized our assessments, can we already consider it as innovative assessments? If we ask students to have their e-Portfolio using various technology tools and apps, can we consider them authentic?

Answer of Dr. Jasper Alontaga

When you talk about innovation, there are different levels and not all innovations are created equally. That's why you need to consider your context and define what innovation means within that framework. It can be as simple as digitizing your assessment and that makes it innovative. But when we pursue the digitalization of assessment to make it innovative, then we have to consider the critical pedagogical practices mentioned earlier.

Question #9

What makes Standardized Tests valid and authentic in terms of skill-based competencies? Aside from developing rubrics, what other objective criteria educators should consider and use?

Answer of Dr. Therese Bustos

I think we need to realize the limitations of standardized testing. There are really limitations, and especially if you are after specific demonstrations of a particular skill. This is why you will most likely have a combination of standardized assessments and also authentic assessments. There have been attempts at trying to have a standardized version of something that's really authentic. But if you look at the responses, for example, if there are four possible responses, and if you look at people being able to guess which of the four would be the correct response, there is a great possibility that a person might be able to give you an answer, but not necessarily be able to demonstrate that. And this is why, personally, I would prefer a combination of standardized and performance-based assessments, just to show us a better picture of what a student can do. I would not try combining authentic, performance-based, and multiple-choice questions in the same assessment. I had put them separate.

Answer of Dr. Jimmy De La Torre

Different types of assessments serve different purposes. Now, the issue of validity, right now is not an intrinsic property that applies to all circumstances. A test could be valid for this particular purpose, but not for another. That's something that you need to establish. You can't just say, oh, I've established my

validity. I can use it for whatever purpose and without renewing or studying it again. That's something that you can't do.

Now, I'm a big proponent also of the use of technology. There are certain circumstances where you have to realize that there's always a trade-off. Let me give you an example. I was consulting for a company in the U.S., Educational Testing Service. That's one of the largest testing companies in the U.S and they spent a million of dollars to develop a more authentic assessment were, okay, you're walking on this. They gave the test and then they also gave multiple choice tests.

And what they found out was the correlation between the score, we're using a more authentic type of assessment, actually correlates very highly, about 0.9. So, in a way, it doesn't capture everything that the more authentic assessment is giving, but at the same time, it's much cheaper to actually administer. So I could imagine there are contexts where you could have a good approximation and save some money and some resources, but there are circumstances where it may not be possible. We really have to sit down with the students and really evaluate the student more individually.

The other aspect is related to, this is actually related to technology also. When they did it, they gave them a film. There's a demo of how things are working.

It's quite nice, but very expensive to really implement. The other aspect is we can make assessments more authentic. I have colleagues in Taiwan, instead of just multiple choice, they're asking, what's the angle of the two lines? Instead of saying, oh, right angles, scaling, or whatever type of answer.

Well, in fact, they said, okay, here's a virtual protractor, and I want you to measure the actual angles. And they actually will be able to go there and give the specific number rather than just memorize what are less than 90 degrees, so on and so forth. The other is they also have, for any chemistry, instead of asking them what will happen when you mix this two, they actually have virtual chemistry lab, where they can mix the chemicals and see what reactions will take place.

To some extent, you can actually make it closer, but not equivalent to what is really authentic in real life. But to some extent, you can approximate it, and you always have to weigh the benefit and cost of doing so. So as I've said, to a large extent, it is a lot of approximation that we're doing.

It's because we are dealing with a lot of constraints, one of which has to do with there are already too many assessments going on. They don't have time to learn because all they do is take assessments almost every week. So we're now reversing it just because we know that we do underscore important assessments, but we spend more time in assessments rather than the actual learning. So, we have to be mindful of that also in the process. Everything will be balanced and we don't lose sight of what's really important.

Question #10

All types of assessment at all levels of the educational grid deal with important issues and consequences. As an important area of educational research, how can CEM evaluate the findings of international assessment studies (such as PISA or TIMMS) and national assessments as well, and utilize them at the local level (school and classroom) to improve the quality of education?

Answer of Ms. Kathryn Tan

One of the things really that we wish we could do given the PISA results is more data disaggregation to understand the sources of errors and be able to understand their areas of difficulty really. I think one starting point we could begin with is to look at where majority of the students are, the level in the PISA assessment and try to match it with local assessments that we have that also assess skills at that level. This process would allow us to further disaggregate where students are in terms of their abilities and proficiencies and really use that as the starting point for intervention because there is great variety in the performance of students in PISA in the Philippines, there's great variety in terms of spread, although majority are really at level one in most of the subjects.

You would anticipate that the appropriate intervention at those different skill levels would be different and therefore the progress towards meeting the minimum proficiency defined by PISA would be more incremental for those students at the lower levels meaning the progress you expect to see from them might be slower, measured with simpler assessments until they get to a level where they can be instructed to move on to the next.

Answer of Ms. Janet Evasco

I think in the context of PISA assessment it is important for us to understand the testing environment we have with PISA and how we do assessment in the classroom and even in the Philippines with our national assessment. These are some areas for us to look into as we go into the next PISA assessment.

SYNTHESIS

Mr. Louie P. Cagasan, Jr., MAPsych
Assistant Professor, UP College of Education

Dr. Jimmy Dela Torre in his keynote speech defined what a balanced assessment is. It should be comprehensive, continuous, efficient, useful, and coherent vertically. What's happening in the national assessment should also be somehow aligned with what's happening in the classroom assessment. Then he also talked about, how to implement it and ensuring that these three aspects, the Assessment Triangle (cognition, observation, and interpretation), are aligned.

Aside from that, there is a growing belief among many educational assessment experts that, if assessment curriculum instruction were more integrally connected, student learning would improve. If we change one aspect of, this diagram, for example, we change curriculum, assessment, and teaching and learning should also change or follow.

Then next, from CEM, Ms. Kathy and Ms. Janet emphasized that assessment can act as a leverage for a quality education. We can use assessment to inform planning and educational decisions, enhance the process of education by providing feedback with support teaching and learning, and substantiate the outputs and outcomes of education. They also emphasize the need for standardized tests to establish reliability and validity evidence to ensure that the data that we're using to implement plans are reliable or trustworthy. CEM, the Center for Educational Measurement, can provide this educational assessment needs.

Meanwhile, Dr. Jasper shared to us the SAMR model for technology integration. There's a natural progression from no technology to transformation. Starting from substitution and augmentation, which is enhancement, then modification and redefinition, which is the transformation.

There are 2 aspects or two dimensions of e-assessment innovations. First is the innovative aspect from traditional to innovative assessments and from paper-based to technology-rich assessment if you want to reach transformational changes in educational assessment. These are the changes that we can do or implement. Dr. Jasper earlier shared to us some insights or tips on how to apply this technological integration.

Finally, Dr. Therese Bustos also shared to us what authentic assessment is, especially, assessment that foster 21st century skills.

She highlighted that authentic assessments must be well thought out and carefully designed, and we need to pilot these assessments whether they're rubric, task, or innovations, and the need to establish validity. She also emphasized that authentic assessments are still proxies.

The definition of authenticity lies in an assessment requiring students to use the same competencies or combinations of knowledge, skills, and attitudes that they need to apply in criterion situation in professional life. The level of authenticity of an assessment is defined by the degree of resemblance to the criterion situation.

Emphasizing on the 21st-century skill, Care and Kim mentioned that measurement of 21st century skill is still in its infancy. Since it's still in its infancy, we will encounter challenges. Most of the challenge would come from our lack of understanding or comprehensive understanding of the nature, development of skills, multidimensionality, and other things about 21st century skill. Some of the questions that we can ask if they are indeed malleable: How to best measure the skills? How are these skills related to other constructs.

Where are we right now? If you read the EDCOM 2 year 1 report, the miseducation. There's a section about measurement of student learning outcomes and there's so many challenges that we have. The sad reality is that we are encountering so many challenges in terms of education. Hopefully, you can read the report.

I would like to highlight the 2 common tensions in educational assessment. The first one is accountability versus instructional improvement. When we start to do assessment sometimes teachers would want to focus on helping students to learn more about this particular topic. However, due to accountability purposes or they need to cover the entire curriculum, they're in a dilemma to choose whether they need to focus on focusing on student or proceeding with the lesson that they're in. Another challenge or tension is the manageability versus desirability now that there are a lot of changes that would need to happen in the Philippine education system. All of those are desirable or we want to aspire or those are things that we want to have. But the question now is, are they manageable to begin with?

According to Hayward, if, manageability and desirability become 2 competing courses, manageability would most likely win. It will go back to the way it was. Hopefully, when we implement these changes, it would always have the criteria of being manageable.

Now the question is, how do we get there? Hopefully, we will be able to address the tensions that weaken the blocking courses and strengthen the facilitative forces and be acquainted and use advanced and robust concepts and tools that was also shared to us by our keynote speaker and plenary speakers.

Just a reminder, when we start to implement changes, there's always the need for space. So, according to Laveault, there should be a space for policy developers and policy implementers to co-regulate or make appropriate adjustments and accommodation. If for example, a policy was implemented, it's possible that there are some changes that needs to be done in the field. Hopefully, co-regulation will happen.

Insanity is doing the same thing over and over again and expecting the different results according to Albert Einstein. When you fail to plan, you are planning to fail. Hopefully, when we try to address some and do assessment innovation, the plan is clear and what we want to have or do is also clear. I hope our road map is clear. When there is no vision, people perish.

Overall, changing or revolutionizing the educational assessment system in the Philippines is a difficult task, but whole learners are holistically developed. They have the competence and skills. They're equipped for work in higher education, and they are lifelong learners. Again, we hope we have the vision that it's *para sa bata* and *para sa bayan*.

CLOSING REMARKS

Ms. Iris Lark H. Dizer, SEdMEv
Overall Chair, CEM 46th Anniversary Virtual Conference

Our CEM Board of Trustees, headed by the Chair, Dr. Luis Sorolla, Jr., CEM President & CEO, Dr. Grace Aguiling-Dalisay, CEM Trustees, CEM Corporators, distinguished guests, partner member schools, friends, and colleagues, good afternoon and happy anniversary!

We've had a truly inspiring and informative morning, filled with thought-provoking discussions and valuable insights on building a balanced assessment system. I hope you are leaving this conference not only inspired but also equipped with practical strategies to address the challenges and opportunities in your respective schools.

To make things more exciting, I will have to add to the challenge. As I reflect on today's sessions, it is clear to me that a balanced assessment system goes beyond collecting data. A balanced assessment system should be supported by a well-crafted plan for data utilization to maximize the impact of assessments. Schools must cultivate a strong culture of data use, where collaboration within and across grade levels and subject areas becomes central to identifying areas of concern and refining practices.

School leaders, among the many responsibilities you hold, establishing and promoting this culture of data use is essential. Both leaders and teachers must understand which assessments are needed, when to administer them, and how to effectively use the results to drive improvement. Yet, as our keynote speaker pointed out earlier, "Selecting appropriate assessment tools and integrating them effectively into teaching and learning remains a challenge." This is where targeted professional development becomes vital. By investing in the continuous growth of our educators, we equip them with the skills to choose and implement assessments that truly enhance learning outcomes.

The story of a provincial school in the Cordillera Administrative Region (CAR) illustrates the critical role of using assessment data to inform and drive educational success. The school conducted a reading comprehension survey to assess the effectiveness of teacher-developed reading intervention materials, which were designed following a CEM-facilitated training program. The data collected revealed remarkable improvements: students at the frustration reading level

dropped from 8 to 0, while independent readers increased from 0 to 7. This is a testament to how purposeful use of assessment data can lead to significant student success. Imagine the impact if we all had similar success stories to share.

As we move forward in improving educational practices, building a balanced assessment system becomes crucial in ensuring that we not only gather data but also understand the full picture of student learning. A well-rounded approach allows for more informed decision-making, ensuring that all aspects of student progress are considered and addressed. However, no matter how well-designed an assessment system is, it becomes ineffective if we fail to use the data meaningfully and foster conversations around it. Data should not just inform us; it should guide us to act, collaborate, and continuously improve.

As I close, I would like to express my deepest gratitude to those who made this event possible. To my dedicated and hard-working teammates, the chairs, and members of the organizing committee—thank you for your all your efforts. To our President and CEO, Dr. Grace Aguiling-Dalisay, for her unwavering support during the planning stages. To all our conference speakers, thank you for generously sharing your expertise, to our moderator for giving an on-point synthesis of the discussions. And to you, our guests, your presence and active participation have made this conference a success—thank you.

A special thanks to our corporate event partners:

- Private Education Assistance Committee
- UPRAISE
- BPI Wealth
- BDO
- Abacus Securities

And our media partners:

- The Philippine Star
- Pilipino Mirror
- The Manila Times

And most importantly, we give praise and give back all the glory to God for this conference.

Thank you once again. God bless us all, and till we meet again.

46th ANNIVERSARY CONFERENCE COMMITTEE

Overall

Chair	Iris Lark H. Dizer
Co-Chair	Jacob Avelino S. Gonzales, Jr.

Programs

Chair	Armi S. Lantano
Co-Chair	Pristine Mae T. Cammayo
Members	Noelyn H. Curutan
	Rinnah Keziah O. Ramirez
	Jerome R. Calicdan

Registration and Invitations

Chair	Rosario P. Faylona
Co-Chair	Cherry Gil M. Otto
Members	Joemalyn G. Rayco
	Manilyn R. Reyes
	Valerie S. Equipador
	Hendrick D. Cabanlit
	Leah B. Parado
	Eleonor B. Belista
	Ronaldo M. Gabriel
	Jennifer R. Palasigue
	Jenaica V. Llanares
	Karen Lisette M. Reyes
	Cyndelle Jiene S. Marfori
	Airene Grace D. Peñones
	Renante S. Mina
	Larry Vincent P. Cagadas
	Luzviminda G. Aguiwas

Production and Technical

Chair	Aileen A. Cartagena
Co-Chair	Aldwin D. Calilong
Members	Ralph Ronald J. Baniqued
	Frances F.M.B. Cardona
	Eric A. Tadeo
	George P. Sibayan

Publicity and Sponsorship

Chair	Jacob Avelino S. Gonzales, Jr.
Co-Chairs	Laarni V. Bundoc (Sponsorship)
	Cristina B. Corpuz (Publicity)
Members	Eleonor B. Belista
	Frances F.M.B. Cardona
	Jenelle C. Dionisio
	Mitzi Faye B. Soriano

Logistics and Venue

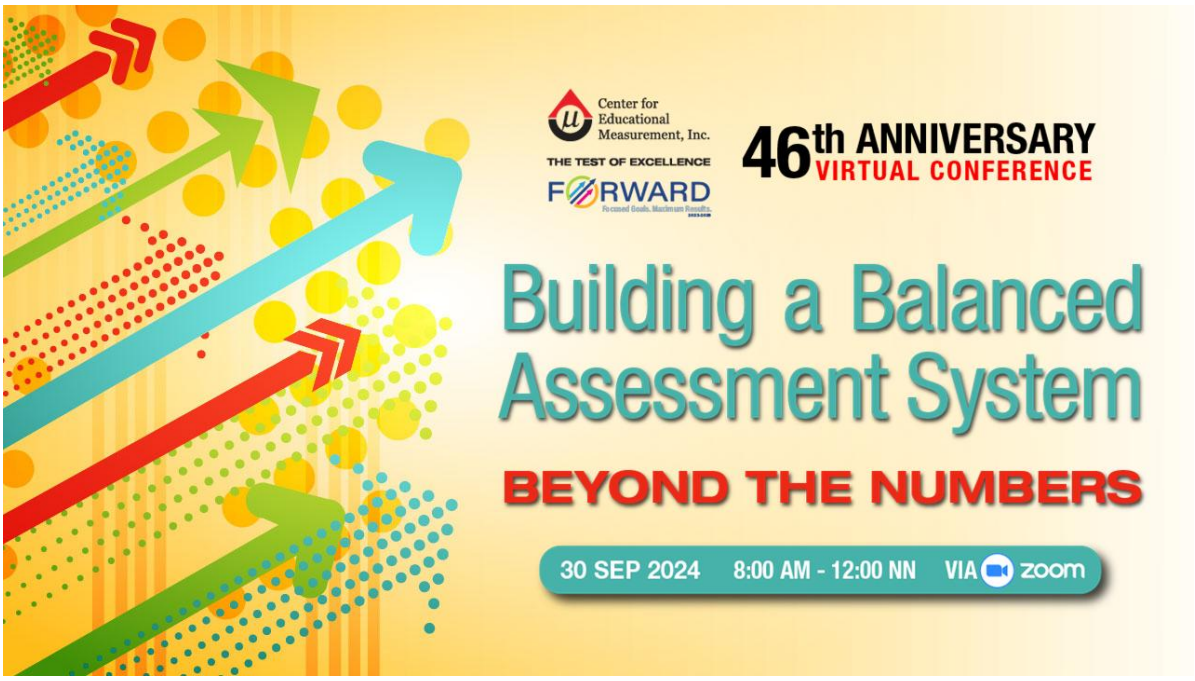
Chair	Maribel A. Montalbo
Co-Chair	Kristine C. Mercado
Members	Jestoni C. Caniedo
	Bernadette S. Escaño
	Anne Grace L. Valdez

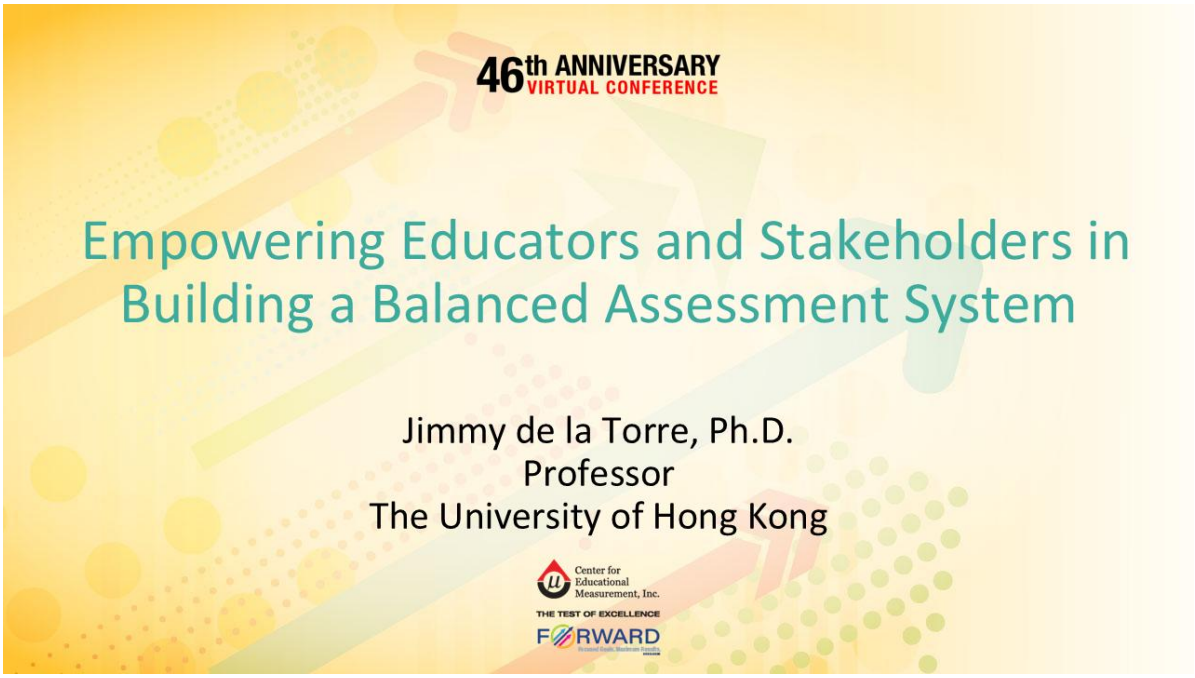
Finance

Chair	Lovelyn R. Santibañez
Co-Chair	Nelsie A. Cabaluna
Members	Raicy M. Notarte
	Criselda E. Sevilla
	Irene Sofia L. Gallivo

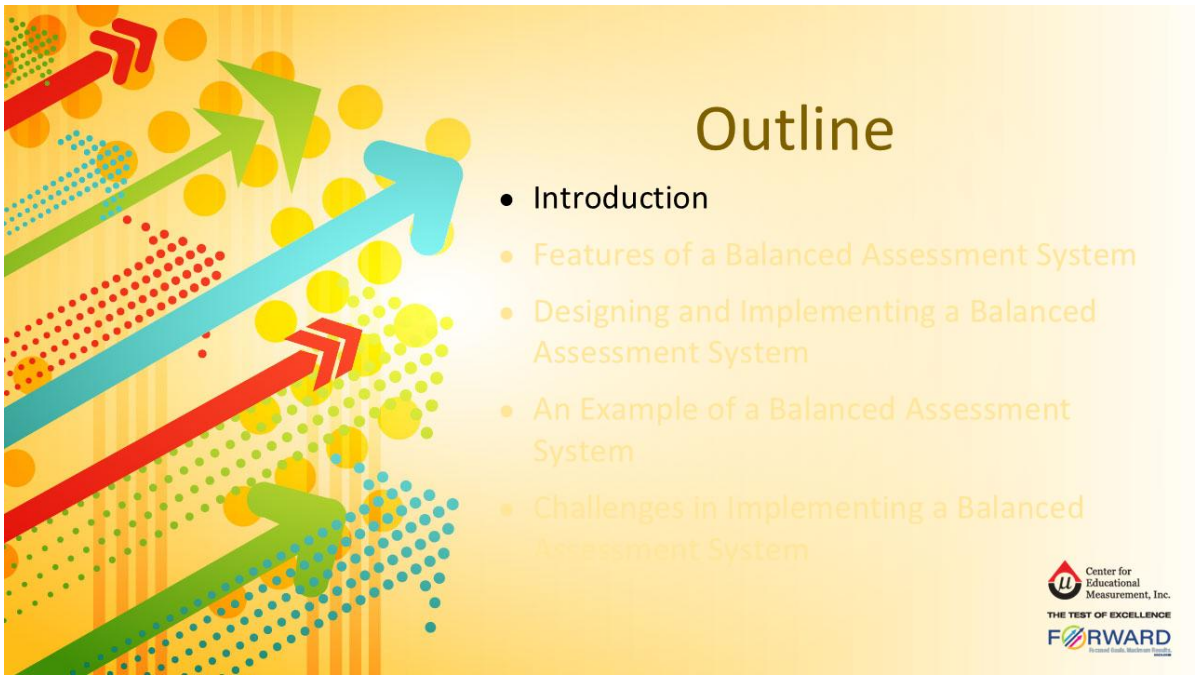
SLIDE PRESENTATIONS OF RESOURCE SPEAKERS

Dr. Jimmy De La Torre





This material is intended for presentation purposes only.
DO NOT CITE OR USE WITHOUT PERMISSION.



Introduction

- The need for a balanced assessment system
  - Assessments are a vital part of education → they give teachers, education leaders, and policy makers important information on what students are learning and where resources can be allocated most effectively
  - Assessments that are administered without considering how they inform teaching, learning, and educational decision making can constrain rather than support student learning and performance
  - When assessments are considered in isolation, how one assessment may interact with other assessments, and how differences across different levels (e.g., classroom, district, national) can impact the effectiveness of an assessment or assessment system can be difficult to understand

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Introduction

- Problems with extant assessments or assessment systems
  - Over-reliance on large-scale standardized tests at the exclusion of other sources of evidence → these assessments are largely summative in nature, are not connected to daily teaching and learning, and fails to provide timely feedback to inform **in-time** instruction and student learning
  - Narrow focus on accountability and grading purposes → these assessments do not adequately consider how assessment could be designed and assessment data used for teaching and learning purposes
  - De-emphasis on formative assessments → traditional (i.e., standardized) assessments do not have fully utilize the formative potential of assessments, which need to be administered during the teaching process to provide **in-time** feedback that can support classroom instruction and student learning

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Introduction

- Problems with extant assessments or assessment systems
  - Lack of Coherence → different assessments are designed in isolation without clear connections to learning objectives or other assessments, thus, leading to assessments that lack coordination and overall coherence
  - In many instances, development and analysis of these assessments are not based on state-of-the-art psychometrics and technologies
- Assessments should **not** be designed and implemented independent of one another
  - Each assessment must be designed to support its intended and specific purpose
  - All assessments must work collectively to address the overall goal of the assessment system

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

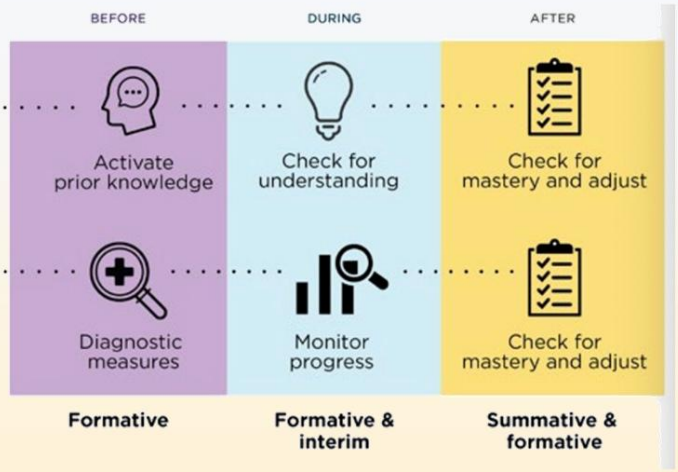
Introduction

- A balanced assessment system is a set of components that interact with each other and function coherently with the intention of improving student learning
- Examples of these components are curriculum, learning theory, instruction, assessment, professional learning, psychometric framework, and implementations
- Such a system allows educators to link assessments through clearly defined learning targets, provide multiple sources of evidence to support decision making, and track progress over time (within a year and across several years)

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Introduction



Sequence and Relationship Between
Formative, Interim, and Summative Assessments

This material is intended for presentation purposes only.
DO NOT CITE OR USE WITHOUT PERMISSION.

Introduction

- A balanced assessment systems should align with the following:
  - Accounts for the varied needs of stakeholders at different levels of the education system
  - Is unified by common targeted goals for student learning
  - Promotes the flow of information across levels in an efficient, intentional manner that informs educational decisions
  - Requires high-quality assessments and high-quality/appropriate use of results (such as assessment literacy)
  - Is comprehensive, continuous, coherent, efficient, and useful
  - Is developed and implemented using the most recent psychometric methods and technological developments

This material is intended for presentation purposes only.
DO NOT CITE OR USE WITHOUT PERMISSION.

Features of a Balanced Assessment System

- A balanced assessment system is comprehensive
  - It allows students to demonstrate their understanding in a variety of ways and reflect the breadth and depth of the target learning goals/content standards
  - It provides a variety of sources of evidence to inform educational decisions at various levels
- The system is continuous
  - It provides consistent and ongoing information about students' learning progress
  - It allows educators to track students' academic growth over time
- The system is efficient
  - Individually, no assessment is redundant
  - Collectively, all assessment are necessary for making education decisions

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Features of a Balanced Assessment System

- The system is coherent
  - The components within an assessment system are interconnected and are compatible with the underlying model of learning
  - Assessment activities are not isolated events, but a well-organized whole that support each other and work together to promote student learning
- Specifically, coherence typically involves the following two aspects
  - Vertical coherence
  - Horizontal coherence

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Features of a Balanced Assessment System

- Vertical coherence refers to the cohesion of the assessments across different educational levels
  - For example, national assessments should be aligned with, say, district- and classroom-level assessments philosophically and purposefully
  - This ensures that the assessments at different levels complement each other and collectively reflect student learning progress
- Horizontal coherence involves alignment between curriculum, instruction, and assessment
  - This means that assessments should be closely aligned with the curriculum and the instruction
  - This ensures that assessments are measuring what students are supposed to be taught and to learn

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Features of a Balanced Assessment System

- For a system to be coherent, assessments need to also be comparable across
  - Different forms or years of administration
  - Different grade levels
- A balanced assessment system is useful
  - It provides the necessary and timely information to make informed decision about specific instruction or education programs it is intended to support
  - It is designed to serve the needs of multiple and diverse stakeholders

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

This material is intended for presentation purposes only.
DO NOT CITE OR USE WITHOUT PERMISSION.

Features of a Balanced Assessment System

Purposes and Uses	Stakeholders and Context
Supporting instruction and learning	Teachers and students within classrooms
Grading and reporting	Teachers/students within classrooms; parents and principals at the school level
Supporting program/curricular evaluation	Principals/teachers at school level; curriculum/assessment leaders at division/district/regional level
Monitoring trends and evaluating equity	Division/district/regional and school leaders; national education leaders and policy makers
Providing data for accountability	Divisions/district/national leaders; national education leaders and policy makers

Common Assessment System Purposes and Uses with Corresponding Stakeholders and Contexts

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

This material is intended for presentation purposes only.
DO NOT CITE OR USE WITHOUT PERMISSION.

Designing and Implementing a Balanced Assessment System

- Designing a balanced assessment system is an iterative process that requires an understanding of how assessments will meet their goals
- It requires
  - Defining a learning theory or framework, and identifying the role assessment will play in that theory or framework
  - Reviewing existing assessment systems or available resources → revising existing systems, eliminating redundant or unnecessary systems, and identifying gaps in information and additional needs
  - Designing a comprehensive plan that will support the assessment needs before selecting existing or developing new assessments to be included in the assessment system

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Designing and Implementing a Balanced Assessment System

- Define the learning theory
  - Educators and experts identify or define a clear theory of learning
  - Theory will guide the direction of and approach to the assessment
- Determine the role of assessment within that theory
  - Understand how students learn
  - Design assessments based on that understanding
- Review existing assessment systems
  - Determine which existing assessments, if any, are effective and supportive of student learning and teacher instruction
  - Identifying and eliminate redundant or unnecessary assessments

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Designing and Implementing a Balanced Assessment System

- Develop a comprehensive plan
  - Ensure that assessment activities are aligned with educational goals and learning theories
  - Articulate the purpose of the assessment, type of assessment, timeline for the assessment, and expected outcomes or improvements
- Selection of assessment tools
  - Assessment tools that are consistent with the comprehensive plan need to be selected or developed
  - They should provide accurate and meaningful information about students' learning status and progress
  - Their validity, reliability, fairness, and applicability need to be established and documented

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Designing and Implementing a Balanced Assessment System

- Integration of assessment practices
  - Ensure that assessment tools and practices are closely integrated with instructional activities → form a system that supports learning and teaching
  - Provide teachers with in-service training or professional development on the proper use of assessment results
  - Allow assessment results to inform adaptation of classroom practices
  - Implement changes to school culture and policies to promote the appropriate use of a balanced assessment system

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Designing and Implementing a Balanced Assessment System: The Assessment Triangle

- Assessment can be viewed as a process of reasoning from evidence
- What all educational assessments have in common is the desire to reason from particular things students say, do, or make, to inferences about what they know or can do more broadly
- An assessment is a tool designed to observe students' behavior and produce data that can be used to draw reasonable inferences about what students know
- The process of collecting evidence to support the type of inferences one wants to draw is referred to as reasoning from evidence

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

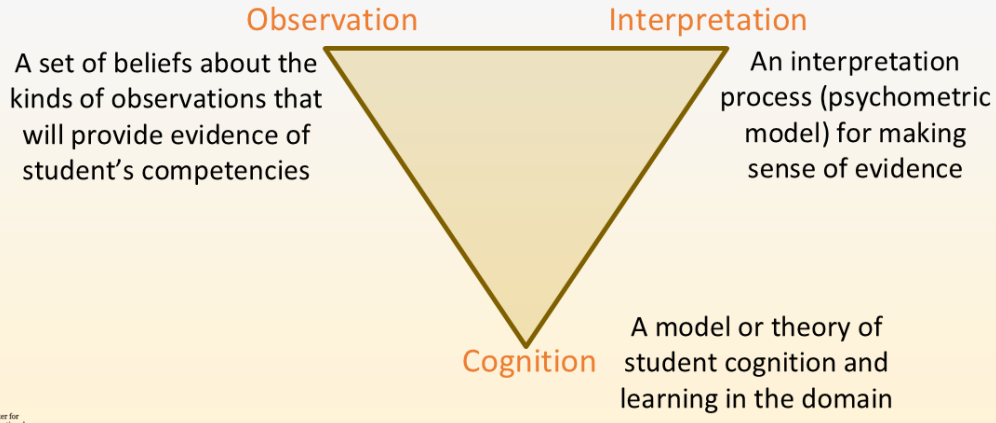
Designing and Implementing a Balanced Assessment System: The Assessment Triangle

- Data become evidence only when one has established their relevance to the conjecture being considered
- Data do not provide their own meaning; their value as evidence can arise only through some interpretational framework
- Educational assessment data become evidence only with respect to conjectures about how students acquire knowledge and skill
- What we believe about the nature of learning will affect the kinds of assessment data sought, and the chain of inferences drawn
- Assessment also depends on which tools are available to make sense of the evidence

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Designing and Implementing a Balanced Assessment System: The Assessment Triangle



The Assessment Triangle

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Designing and Implementing a Balanced Assessment System: The Assessment Triangle

- Classical test theory (CTT) and item response theory (IRT) are the main psychometric frameworks used for interpreting assessment evidence → CTT works at the test level, whereas IRT at the item level
- Cognitive diagnosis models (CDM) in an emerging psychometric framework that can be used for interpreting assessment evidence
- Item-level models (i.e., IRT and CDM) are more flexible and powerful, hence, are used in more modern testing programs
- IRT and CDM serve difference purposes → IRT for is used with summative assessments, whereas CDM with formative assessments

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Fundamental Difference between IRT and CDM

- IRT: performance is based on a unidimensional continuous latent trait θ
- Students with higher latent traits have higher probabilities of answering the question correctly
- CDM: performance is based on a binary latent attribute vector $\boldsymbol{\alpha} = (\alpha_1,\ldots,\alpha_K)$
- Successful performance on the task requires a series of successful implementations of the skills specified for the task

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

An Example of a Proportional Reasoning Test Item

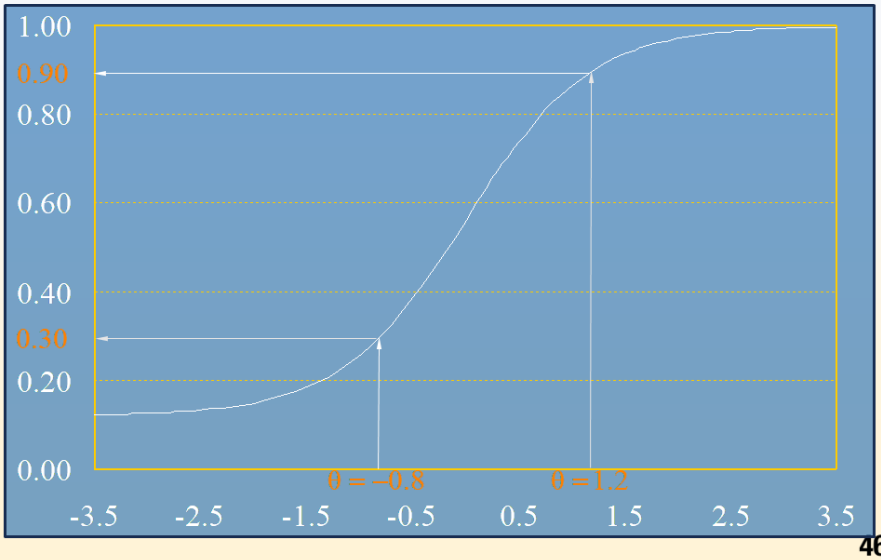
Leon works in a hardware store and is arranging the wrenches on a wall display. He is missing a wrench of size between 5/8 inch and 3/4 inch. Which size wrench is he missing?

a) 6/8 inch
b) 7/8 inch
c) 9/16 inch
d) 11/16 inch

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

A Key IRT Assumption



Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

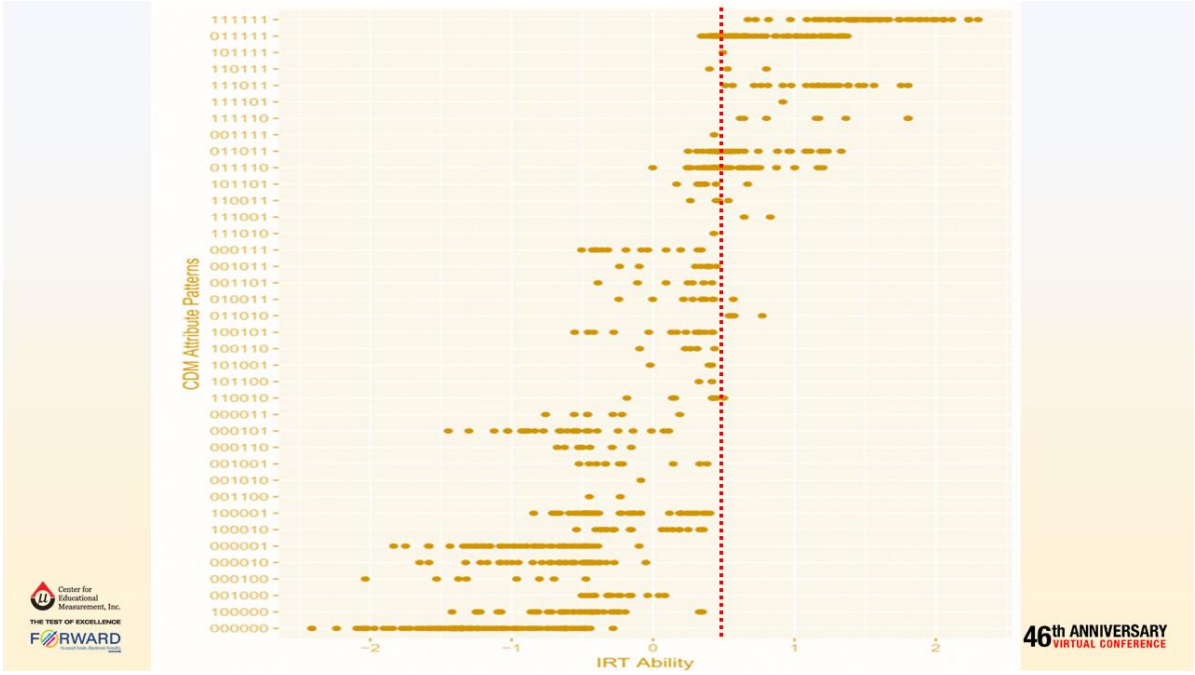
46th ANNIVERSARY VIRTUAL CONFERENCE

An Example of a Proportional Reasoning Test Item: Secondary Students

1. Prerequisite skills
2. Comparing and ordering fractions
3. Constructing ratios and proportions
4. Identifying a multiplicative relationship
5. Determining proportional relationships
6. Applying algorithms

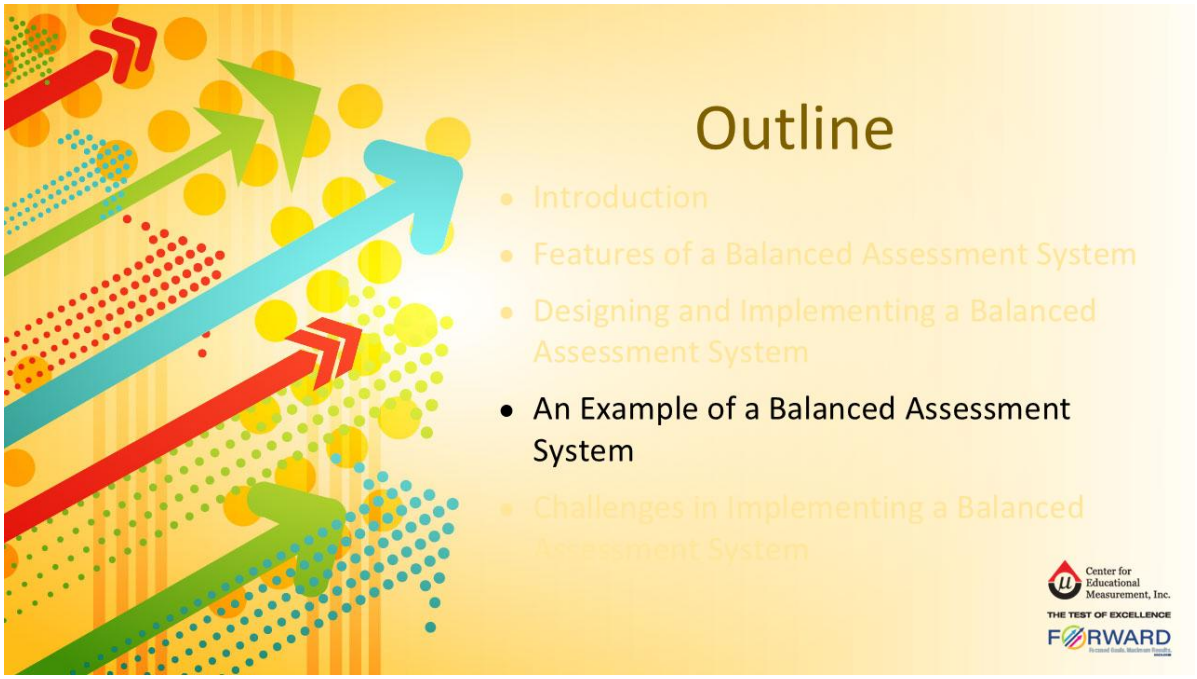
Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE



Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE



This material is intended for presentation purposes only.
DO NOT CITE OR USE WITHOUT PERMISSION.

An Example of a Balanced Assessment System

- The English MAP® Reading Fluency™ is a product of the Northwest Evaluation Association (NWEA)
- It is an adaptive online assessment of early literacy that supports students on their path to reading comprehension by assessing and helping to improve both oral reading fluency and foundational reading skills
- The test adapts in real-time and presents easier or more difficult passages and items depending on student performance
- All spoken-response tasks are scored via automated speech recognition (ASR) software

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

An Example of a Balanced Assessment System

- It is comprehensive
  - Teachers can choose from various test forms, the Adaptive Oral Reading form is the default
  - Administration procedures can vary to accommodate a variety of student and educator needs
  - MAP Reading Fluency is designed to offer one source of data for comparing a student's reading fluency to a general grade-level expectation
- It is continuous
  - Students generally take a 20-minute benchmark assessment three times a year in fall, winter, and spring
  - The assessment is automatically scored, with results appearing in the educator reporting site

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

An Example of a Balanced Assessment System

- It is efficient
  - It is automatically scored and generates actionable data about students' reading skills and instructional needs, including a universal screening outcome
  - It provides immediate and rich information to allow for individualized reading instruction
- It is useful
  - Proficiency relative to grade-level expectations and individualized literacy profile is provided for each student
  - It also makes recommendation of possible next steps

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

An Example of a Balanced Assessment System

- It is coherent
  - If the student is not ready to read passages, they are presented with a series of measures that assess foundational reading skills
  - Teachers can also play back the audio recording for further evaluation and design their own intervention
  - It uses a stable scale that accurately measures student performance regardless of age, grades, or grade level → the scale can track how much students have grown between tests

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

An Example of a Balanced Assessment System

- The test design is based on the Simple View of Reading model (Gough & Tunmer, 1986), a research-validated model of reading development
- It aligns all the decoding, language comprehension, and fluency measures to the Common Core State Standards (CCSS; National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010)
- It offers two broad sets of content: oral reading fluency and assessment of foundational reading skills. Students are routed to one or the other of these sets of content based on their performance on a Sentence Reading Fluency measure
- Below are the part of the domains and measures that comprise MAP Reading Fluency test

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

An Example of a Balanced Assessment System

Domain	Measure	Code	Duration
Foundational Skills			
Phonological Awareness	Rhyme Completion	030	2 minutes, speeded
	Counting Syllables	017	1 minute, speeded
	Onset-Rime Blending	018	1 minute, speeded
	Initial Sound Matching	001	2 minutes, speeded
	Blending Phonemes	019	1 minute, speeded
	Phoneme Counting	020	1 minute, speeded
	Phoneme Addition/Deletion	021	2 minutes, speeded
	Phoneme Substitution	022	2 minutes, speeded
Phonics & Word Recognition*	Letter Knowledge	002	1 minute, speeded
	Letter-Sound Fluency	003	1 minute, speeded
	Build Words: One Letter	024	1 minute, speeded
	Word Families: Initial Letter	023	1 minute, speeded
	Decoding: CVC	007	1 minute, speeded
	Build Words: CVC	025	2 minutes, speeded
	Decoding: Single Syllable	027	1 minute, speeded

Assessed Domains and Measures of English MAP Reading Fluency (Partial)

An Example of a Balanced Assessment System

- For each component of foundational literacy included in the assessment, NWEA content specialists and external experts reviewed the relevant research literature and the relevant academic standards and progressions
- After identifying critical domains and components, they determined the evidence necessary to demonstrate the knowledge and skills represented in each component
- Content specialists created item templates for each measure to ensure consistency across items in content scope, context, cognitive complexity, item format, graphics, and audio style
- At the item template level, the approach and phrasing of the stem was determined and reviewed for best item construction practices

An Example of a Balanced Assessment System

- IRT is used to place items and persons onto the same scale
- It allows student scores to be readily compared even when students have taken different sets of items on the same scale
- The IRT calibrations are intended to provide an item bank for future scoring of students, support the creation of scaled scores for each domain, and provide an historical, longitudinal dataset
- NWEA partnered with LanguaMetrics to develop the speech scoring engine that scores the Oral Reading measures with words correct per minute (WCPM) reported outcomes
- Reports can be generated at various levels (e.g., student, class, school, district)

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

References

- Peng, F., Xue, K., Hall, C., & Newburn, J. (2024), English MAP Reading Fluency Technical Report. https://www.nwea.org/uploads/English-MAP-Reading-Fluency-Technical-Report-2024-08-26.pdf#page25
- Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. Remedial and Special Education, 7(1), 6-10. https://doi.org/10.1177/074193258600700104

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

An Example of a Balanced Assessment System

Rhyme Completion	Listen to four word choices. Given the first two words in a rhyming set, choose the word that completes the trio of rhyming words.	

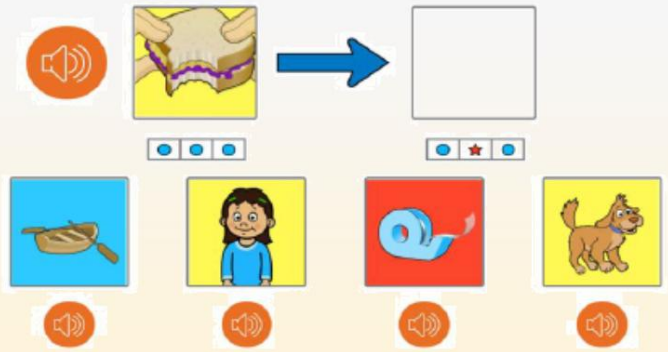
Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

Sample Item – Rhyme Completion

46th ANNIVERSARY VIRTUAL CONFERENCE

This material is intended for presentation purposes only.
DO NOT CITE OR USE WITHOUT PERMISSION.

An Example of a Balanced Assessment System

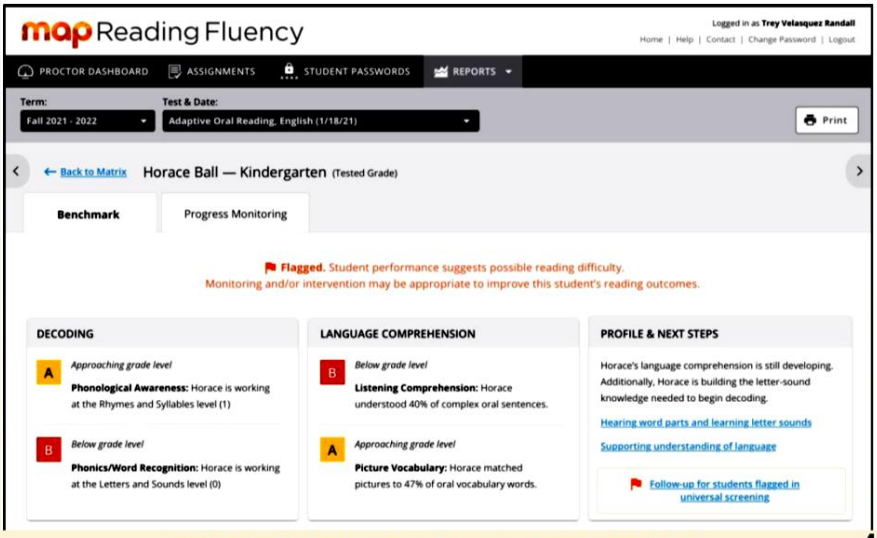
Phoneme Substitution	Listen to a word aloud. Change the middle sound and choose the new word.	

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

Sample Item – Phoneme Substitution

46th ANNIVERSARY VIRTUAL CONFERENCE

An Example of a Balanced Assessment System



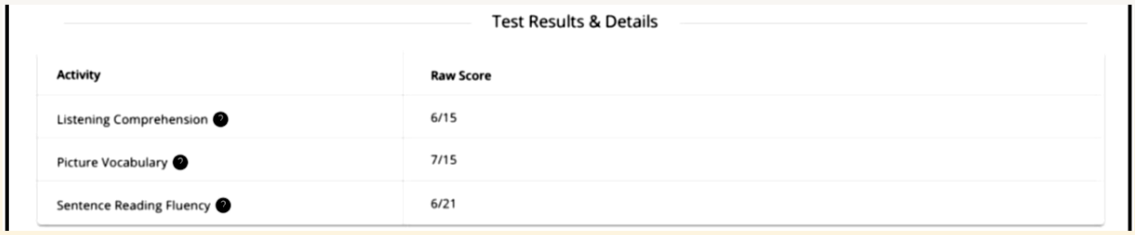
Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

Sample Individual Student Report – Foundational Skills

46th ANNIVERSARY VIRTUAL CONFERENCE

This material is intended for presentation purposes only.
DO NOT CITE OR USE WITHOUT PERMISSION.

An Example of a Balanced Assessment System



Test Results & Details

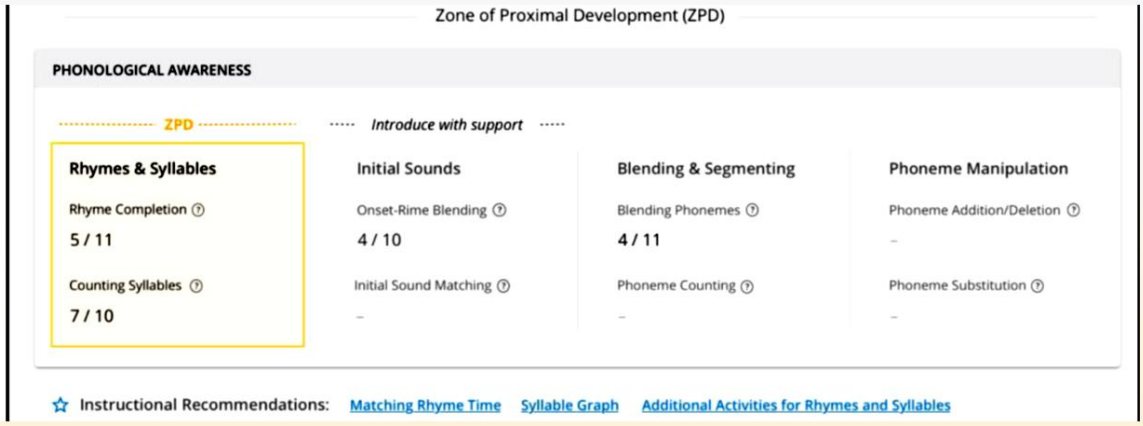
Activity	Raw Score
Listening Comprehension	6/15
Picture Vocabulary	7/15
Sentence Reading Fluency	6/21

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

Sample Individual Student Report – Foundational Skills

46th ANNIVERSARY VIRTUAL CONFERENCE

An Example of a Balanced Assessment System



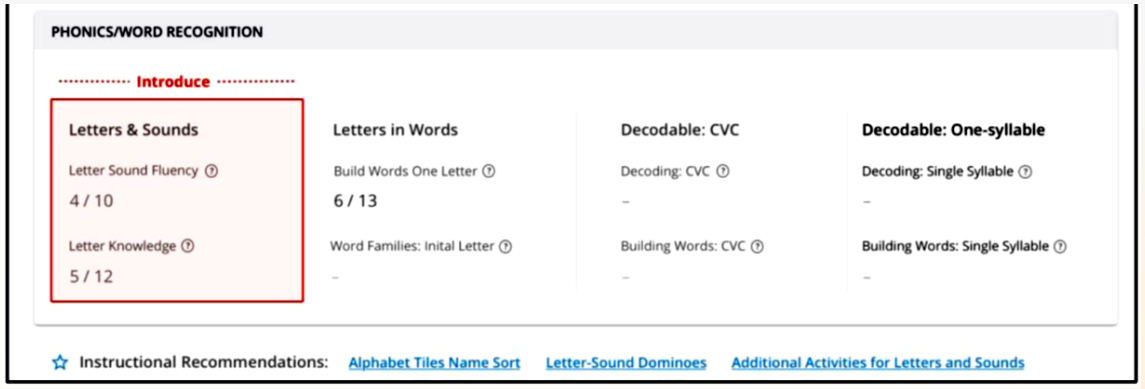
Center for Educational Measurement, Inc. THE TEST OF EXCELLENCE FORWARD

Sample Individual Student Report – Foundational Skills

46th ANNIVERSARY VIRTUAL CONFERENCE

This material is intended for presentation purposes only.
DO NOT CITE OR USE WITHOUT PERMISSION.

An Example of a Balanced Assessment System



Center for Educational Measurement, Inc. THE TEST OF EXCELLENCE FORWARD

Sample Individual Student Report – Foundational Skills

46th ANNIVERSARY VIRTUAL CONFERENCE

An Example of a Balanced Assessment System

Grade	Fall 2020 N	Fall 2020 SS Mean	Fall 2020 SS SD	Fall 2020 SEM Mean	Fall 2020 SEM SD	Winter 2021 N	Winter 2021 SS Mean	Winter 2021 SS SD	Winter 2021 SEM Mean	Winter 2021 SEM SD
Phonological Awareness										
K	75,412	490.94	9.25	2.70	0.93	70,260	494.36	9.44	2.67	0.88
1	91,393	498.90	9.16	2.84	1.01	72,632	501.34	9.58	2.95	1.10
2	61,536	503.57	8.94	3.14	1.21	35,068	503.53	9.32	3.13	1.23
3	13,442	504.10	8.89	3.24	1.27	9,175	504.93	9.14	3.29	1.32
4	844	506.62	9.19	3.27	1.26	547	507.42	9.40	3.36	1.41
5	337	507.51	10.01	3.35	1.31	320	508.26	8.71	3.44	1.49

Center for Educational Measurement, Inc. THE TEST OF EXCELLENCE FORWARD

Descriptive Statistics for Scale Scores and SEMS – Foundational Skills

46th ANNIVERSARY VIRTUAL CONFERENCE

This material is intended for presentation purposes only.
DO NOT CITE OR USE WITHOUT PERMISSION.

Outline

- Introduction
- Features of a Balanced Assessment System
- Designing and Implementing a Balanced Assessment System
- An Example of a Balanced Assessment System
- Challenges in Implementing a Balanced Assessment System

Center for Educational Measurement, Inc. THE TEST OF EXCELLENCE FORWARD

Challenges in Implementing a Balanced Assessment System

- Influence of policy and political boundaries
  - The design and implementation of assessment systems are influenced by different political entities controlling different levels of the education system and corresponding assessments
  - This can lead to inconsistencies in or incoherent assessment systems
- High demands on assessment literacy and design difficulties
  - There is a lack of understanding of how to effectively integrate formative, interim and summative assessment practices
  - Selecting appropriate assessment tools and integrating them effectively into teaching and learning remains a challenge.

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Challenges in Implementing a Balanced Assessment System

- Problems of stability and continuity
  - Assessment systems need to be stable, but frequent changes in policy may lead to a lack of consistency in the assessment system over time
  - This can affect the validity and reliability of assessments
- Neglect of curriculum and learning
  - Designing an assessment system requires considering the needs of multiple parties → insufficient attention may be paid to the needs of curriculum and learning
  - This can lead to a disconnect between the assessment system and the learning pathways and needs of students

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Challenges in Implementing a Balanced Assessment System

- Lack of expertise in modern psychometrics
  - Linking test scores across different time points within a year and across multiple years require the use of proper scaling procedures
  - Instead of relying solely on classical test theory, different paradigms (i.e., item response theory, cognitive diagnosis modeling) may be needed for different types of tests or test purposes
  - Implementing adaptive testing can shorten testing time without sacrificing test reliability
  - Reporting of test scores in a timely manner require moving away from paper-and-pencil tests and/or human graders
  - Individuals with proper training in modern psychometrics can be scarce

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Challenges in Implementing a Balanced Assessment System

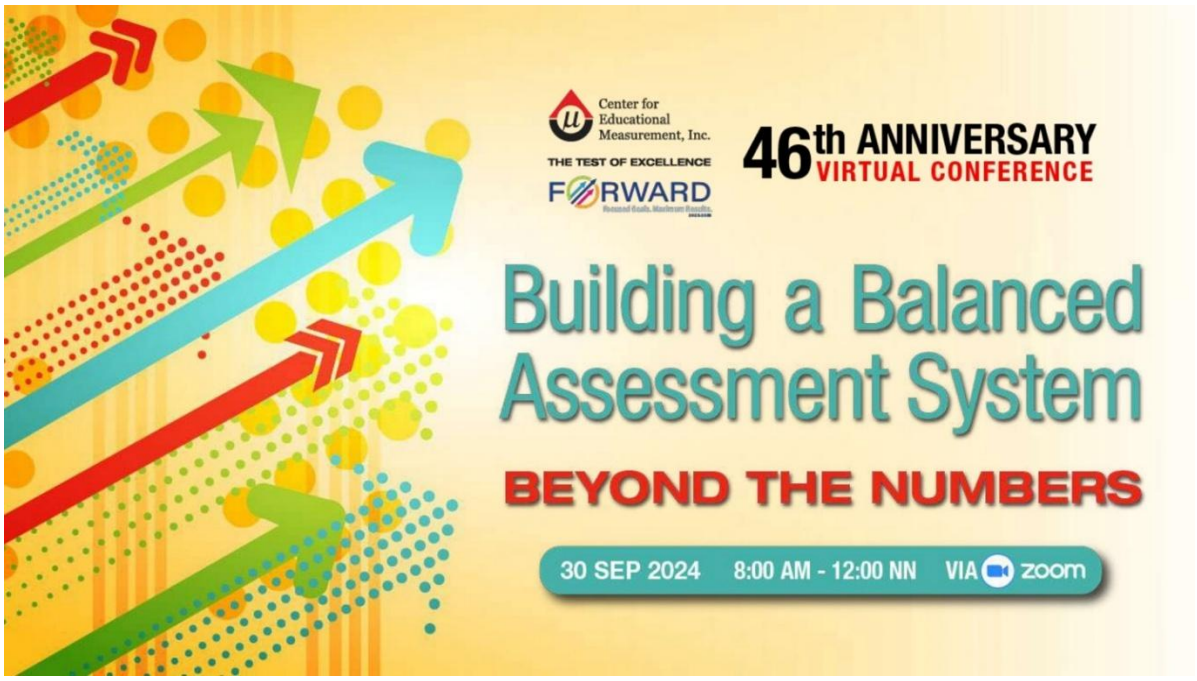
- Difficulties in integrating relevant technology
  - Switching to computer-based testing, which can use alternative test formats to tap skills and abilities not accessible to multiple-choice questions (e.g., 21st century skills), has a high associated cost
  - Computer-based, as well as computerized adaptive testing, requires investing in and building the necessary technological infrastructure
  - Ensuring device access, connectivity, bandwidth sufficiency, and platform compatibility cannot be always guaranteed across the board
  - Updating teachers' pedagogical knowledge and skills to incorporate new technology in their classroom practices requires deliberate effort and time

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

This material is intended for presentation purposes only.
DO NOT CITE OR USE WITHOUT PERMISSION.

Ms. Kathryn M. Tan, MAPsych



Understanding Quality in Education

Notions of Quality Education

Quality education has different interpretations and perspectives, generally referring to judgments about the inputs, processes, outputs, and outcomes of education.

Inputs and Processes

When viewed as inputs or processes, quality education pertains to characteristics of teachers, students, facilities, curriculum, and resources, as well as interactions among these educational elements.

Outputs and Outcomes

As outputs or outcomes, quality education emphasizes desired results like completion rates or achievement of learning outcomes and skills, distinguishing short-term outputs from longer-term effects such as employment or earnings.

THE TEST OF EXCELLENCE FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Role of Assessment in Education

What is Assessment

A systematic method to obtain information and draw inferences about characteristics of people, objects, or programs.

How to do Assessment

Includes various information-gathering methods serving different educational purposes.

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

What are Standardized Tests

Standard Procedures

Standardized tests are designed to have consistent / standard procedures for content, administration, scoring, and reporting of results.

Standardization Enables

Standardization allows for the objective description and comparison of individuals or programs.

Variations in Standardized Tests

Tests can vary by function (e.g., achievement vs. career guidance), scope (national vs. regional), interpretation (norm-referenced vs. criterion-referenced), and granularity of reporting (individual vs. group results).

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Role of Assessment in Education

What is Assessment

A systematic method to obtain information and draw inferences about characteristics of people, objects, or programs.

How to do Assessment

Includes various information-gathering methods serving different educational purposes.

Why do Assessment

Supports decision-making in a variety of educational contexts

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Assessment supports decision-making in various educational contexts

- Teaching and learning
- Guidance and counselling
- Admissions and scholarships
- Program planning and evaluation
- Educational research

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Assessment in Teaching and Learning

Formative Assessment

- Formative assessment is the appraisal of student performance that aims to give feedback for improving ongoing instruction or for directing student effort as they work toward the learning targets.
- It is an integral part of the instructional process, conducted throughout the learning period.
- Formative assessment practices include clarifying learning goals, giving feedback, adjusting instruction based on identified learning needs, and doing student peer- or self-assessment.

Summative Assessment

- Summative assessment is the 'final' appraisal of what students know and can do.
- It occurs at the end of a period of instruction, used for certifying learning, influencing promotion or graduation decisions.
- Well-crafted summative assessments can support learning by facilitating knowledge consolidation and strategy rehearsal.
- Taking tests can strengthen memory retention and reduce forgetting rates.

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Assessment in Guidance and Counselling

Key Points

- Assessments are essential tools used by guidance counselors to help students navigate their educational, career, and life choices.
- These assessments aid in addressing students' uncertainties related to their strengths, weaknesses, interests, and aptitudes.
- By providing insight into students' capabilities, assessments facilitate a thoughtful process of self-analysis and self-reflection.
- This process enables students to better understand themselves and explore various options aligned with their personal aspirations.
- Ultimately, assessments guide students' decisions for the future.

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Assessment in Admissions and Scholarships

Key Points

- The purpose of admission testing is to maximize the selection of students who have a high likelihood of performing well in a particular program of study.
- Scholarship tests function similarly to admission tests, selecting candidates with the highest performance potential, ensuring expanded access to educational opportunities.
- When applicants come from varied backgrounds, it becomes critical to document their competencies and abilities in a comparable way, to facilitate the selection process.

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Assessment in Program Planning and Evaluation

Formative Decisions

- Assessments can be used to describe and appraise programs as well as people.
- They support formative decisions by identifying needs or gaps that require intervention.
- Assessment data is crucial for selecting, designing, or planning interventions and monitoring their implementation.
- Formative assessments inform adjustments and improvements in programs and services.

Summative Decisions

- Summative assessments at the end of program implementation help evaluate the effectiveness or worth of the conducted interventions.
- These evaluations can have high-stakes consequences
- They aid in making decisions regarding the continuation, expansion or closure of particular programs.

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Assessment in Educational Research

Key Points

- Assessments provide measurable data crucial for understanding educational processes and outcomes.
- They focus on variables like student characteristics, school-level factors, achievement levels, socio-emotional variables, and instructional variables.
- Insights from assessments help deepen our understanding of learning processes and the factors that enhance or impede them.
- These insights contribute to developing effective interventions and educational practices to address educational challenges.

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Linking Assessment and Quality Education

Assessment *informs* planning and decisions related to improving **inputs of education**, as in selection and resource allocation decisions.

Assessment *enhances* the **processes of education** by providing feedback that supports teaching and learning or that informs learners' educational or career choices.

Assessment *substantiates* the **outputs and outcomes of education**, documenting achievement of educational objectives and broadening access to educational opportunities.

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Value Added by Standardized Tests

- Extensive review and trialling
- Comparative data for benchmarking
- Validate knowledge and skills
  - Test of transfer
  - Assess learning of intended curriculum
  - Efficient way of presenting evidence of qualifications
- Equalize access to educational opportunities, particularly for those whose capabilities or aptitude might otherwise go unrecognized
- Systematically identify, understand potential and interests

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Assessment Solutions Offered by CEM

Tests of Competencies in Basic Education

- K to 12 Achievement Tests
- Reading Tests
- Pagbasa Tests

Tests for Career Decision-Making

- CEM Career Interest Profiler (CEM Profiler)
- Philippine Aptitude Classification Test (PACT)

Tests for Admissions to Higher Education

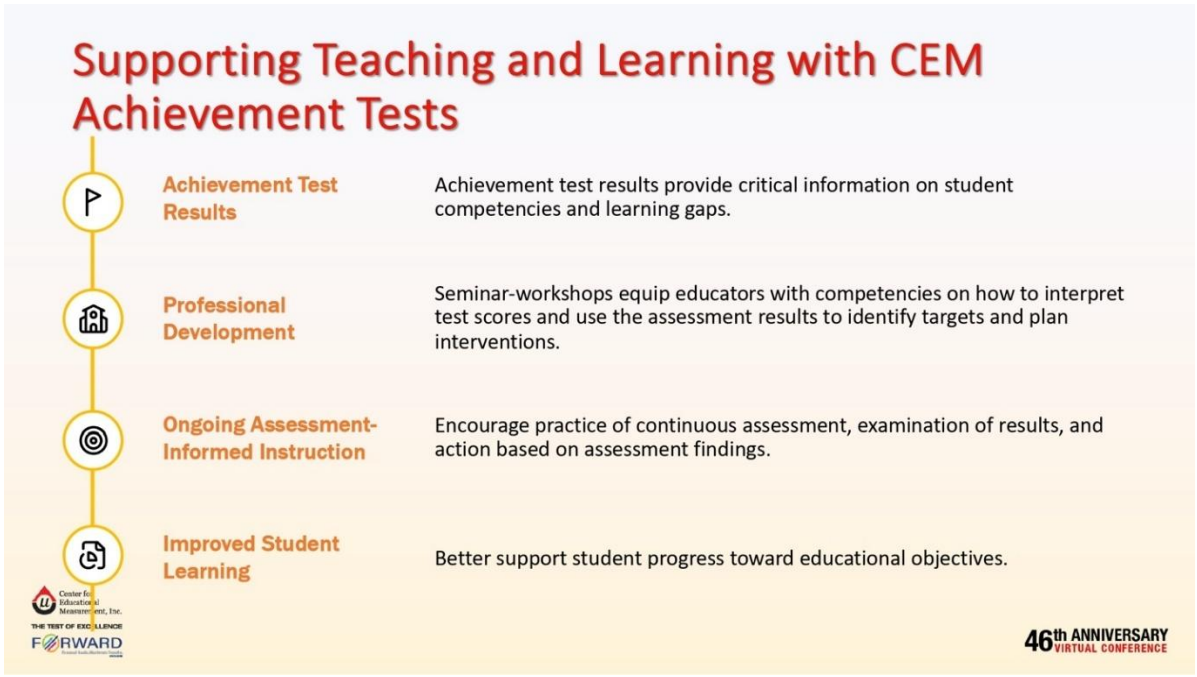
- Elementary Level Competency Measure (ELCOM)
- Readiness Test for Senior High (RTSH)
- Readiness Test for Colleges and Universities (RTCU)
- Nursing Aptitude Test (NAT)
- Allied Medical Courses Admission Test (AMCAT)
- Law School Qualifying Test (LSQT)
- Graduate Level Test (GLT)

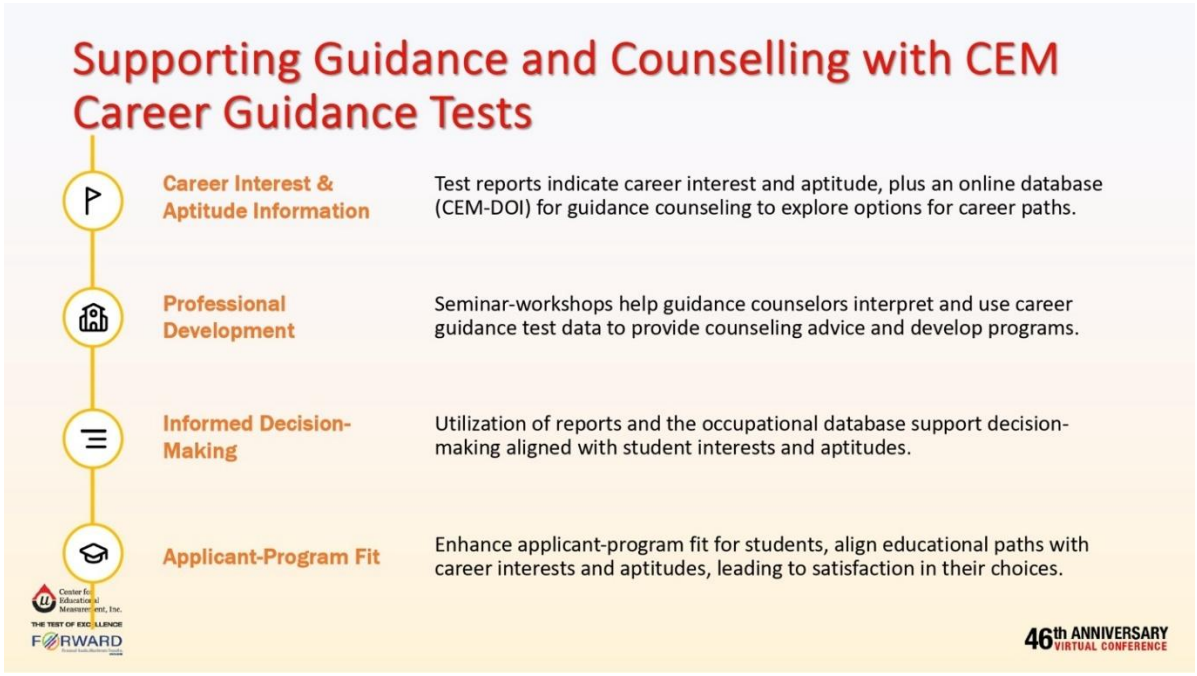
Special Purpose Assessments

- Philippine Aptitude Test for Teachers (PATT)
- Advance Placement Tests (APT for English & Mathematics)
- Panukat ng Pagkataong Pilipino (PPP) English Version by Dr. Annadaisy J. Carlota

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE


Supporting Teaching and Learning with CEM Achievement Tests
Achievement Test Results
Achievement test results provide critical information on student competencies and learning gaps.
Professional Development
Seminar-workshops equip educators with competencies on how to interpret test scores and use the assessment results to identify targets and plan interventions.
Ongoing Assessment-Informed Instruction
Encourage practice of continuous assessment, examination of results, and action based on assessment findings.
Improved Student Learning
Better support student progress toward educational objectives.
46th ANNIVERSARY VIRTUAL CONFERENCE


Supporting Guidance and Counselling with CEM Career Guidance Tests
Career Interest & Aptitude Information
Test reports indicate career interest and aptitude, plus an online database (CEM-DOI) for guidance counseling to explore options for career paths.
Professional Development
Seminar-workshops help guidance counselors interpret and use career guidance test data to provide counseling advice and develop programs.
Informed Decision-Making
Utilization of reports and the occupational database support decision-making aligned with student interests and aptitudes.
Applicant-Program Fit
Enhance applicant-program fit for students, align educational paths with career interests and aptitudes, leading to satisfaction in their choices.
46th ANNIVERSARY VIRTUAL CONFERENCE

Supporting Admissions and Scholarship with CEM Tests for Admission to Higher Education

Information on Aptitude

Test reports indicating aptitude, measuring relevant competencies.

Reasoned Selection

Selection is rationalized based on assessment results, equalizing access to educational opportunities.

Enhanced Fit

Enhance the fit between applicants and programs, leading to better academic performance and program completion.

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Assessment serves as a critical lever for quality education

Assessment informs planning and decisions related to improving inputs, processes, outputs, and outcomes in education. It is not merely an auditing mechanism but can positively influence development when well-designed. Ultimately, through these decision-making pathways, assessment serves as a lever for quality education, enhancing student outcomes and supporting learners effectively.

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

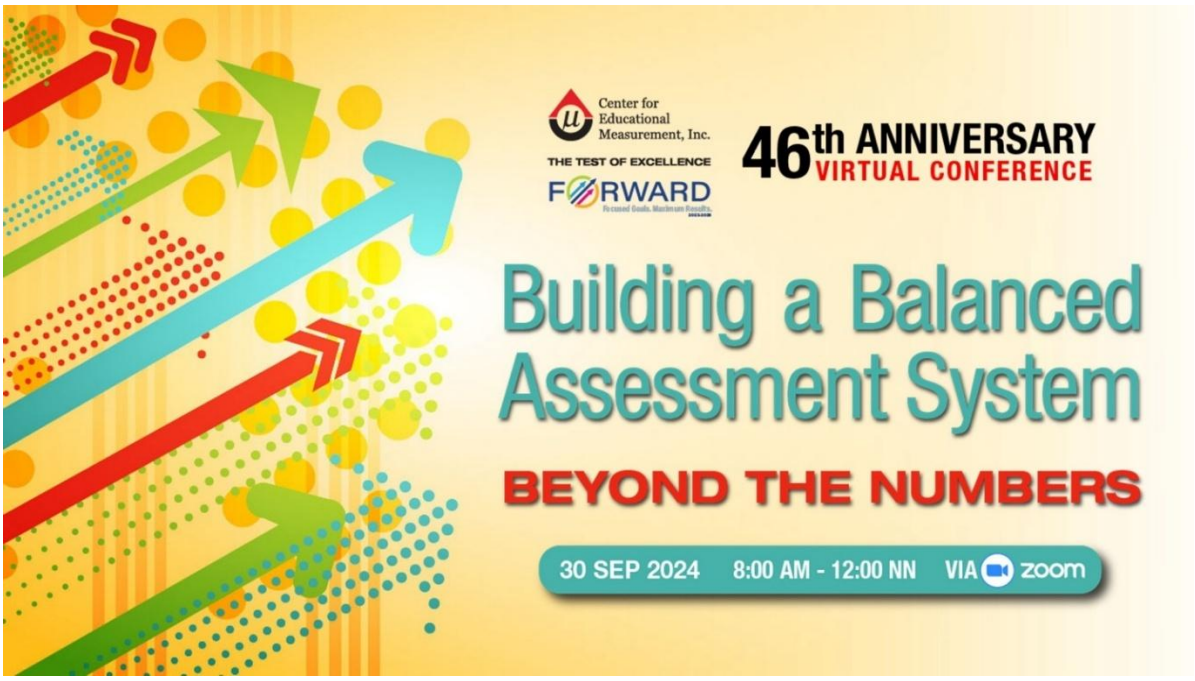
References

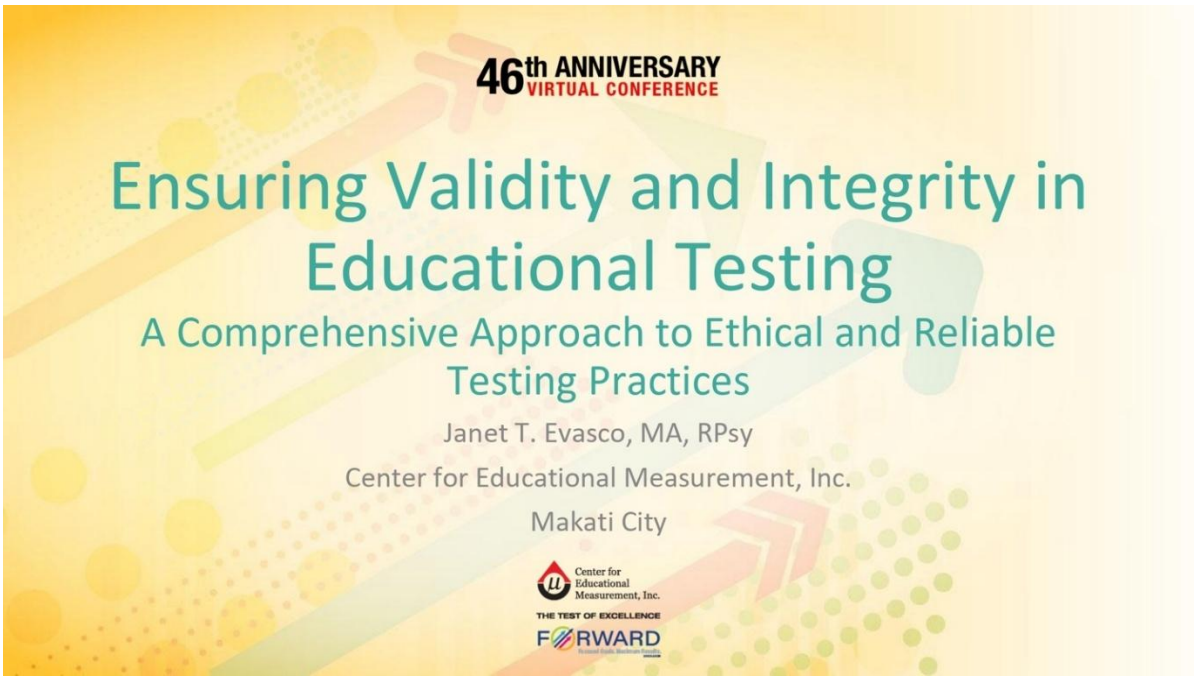
- Adams, D. (1993). Defining educational quality. Improving Educational Quality Project Publication #1: Biennial Report. Arlington, VA: Institute for International Research.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (Eds.). (2014). Standards for educational and psychological testing. American Educational Research Association.
- Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education, Principles, Policy, & Practice,* 18 (1): 5-25.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Dylan, W. (2003). *Assessment for learning.* New York: Open University Press, McGraw-Hill Education.
- Bloom, B.S., Hastings, J.T., Madaus, G.F. (1971). *Handbook on formative and summative evaluation of student learning.* New York: McGraw-Hill Book Company.
- Chatterji, M. (2003). *Designing and using tools for educational assessment.* Boston: Allyn and Bacon
- Earl, L. (2003) Assessment as Learning: Using Classroom Assessment to Maximise Student Learning. Thousand Oaks, CA, Corwin Press.
- Puma, M. & Raphael, J. (2001). Evaluating standards-based professional development for teachers: A handbook for practitioners. Prepared by The Urban Institute for U.S. Department of Education, Planning and Evaluation Service. https://www.urban.org/research/publication/evaluating-standards-based-professional-development-teachers
- Sadler, D.R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18,* 119-144.

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Ms. Janet T. Evasco, MA, RPsy





46th ANNIVERSARY VIRTUAL CONFERENCE

The Critical Role of Validity in Test Quality

- **Validity ensures that test scores accurately reflect their intended purpose.**
- **Evidence of validity is key to confirming that score interpretations are correct.**
- **Standard processes must be followed to maintain the validity of test outcomes.**

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

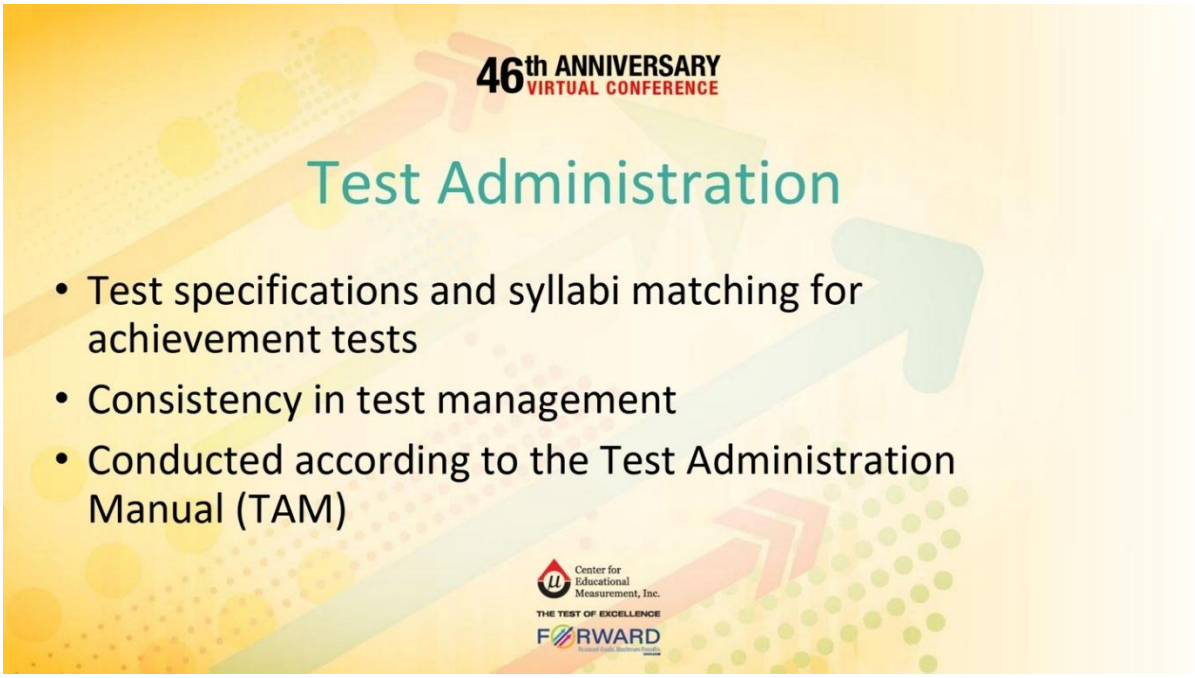
46th ANNIVERSARY VIRTUAL CONFERENCE

Processes to Ensure Test Integrity

1. Test Development
2. Test Administration
3. Training and Support of Testing Staff
4. Creating a Comfortable Testing Environment for Examinees
5. Preventing Fraudulent Means of Obtaining Scores
6. Scoring, Data Security and Confidentiality

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD


46th ANNIVERSARY
VIRTUAL CONFERENCE
Test Development
• Team is composed of test development specialists and subject area experts
• Items or questions are put through multiple reviews and field tryouts to meet the highest standards for quality in the testing industry
Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD


46th ANNIVERSARY
VIRTUAL CONFERENCE
Test Administration
• Test specifications and syllabi matching for achievement tests
• Consistency in test management
• Conducted according to the Test Administration Manual (TAM)
Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Training and Support of Testing Staff

- Comprehensive training program
- Ongoing support
- Technology integration

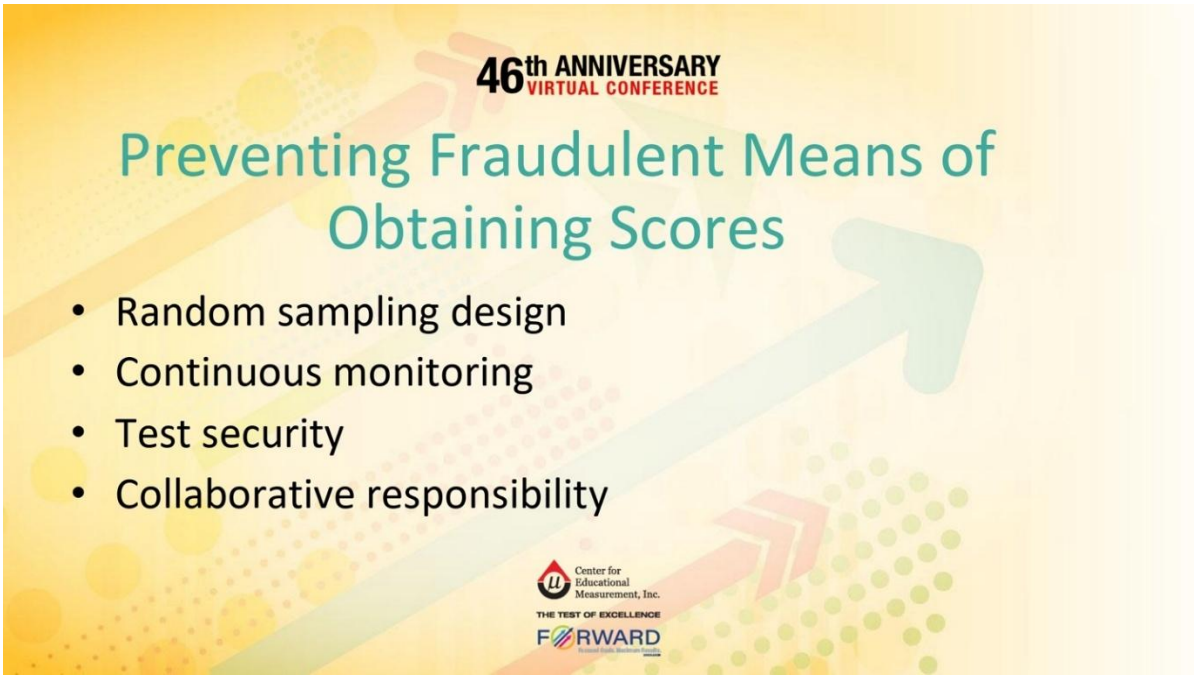
Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

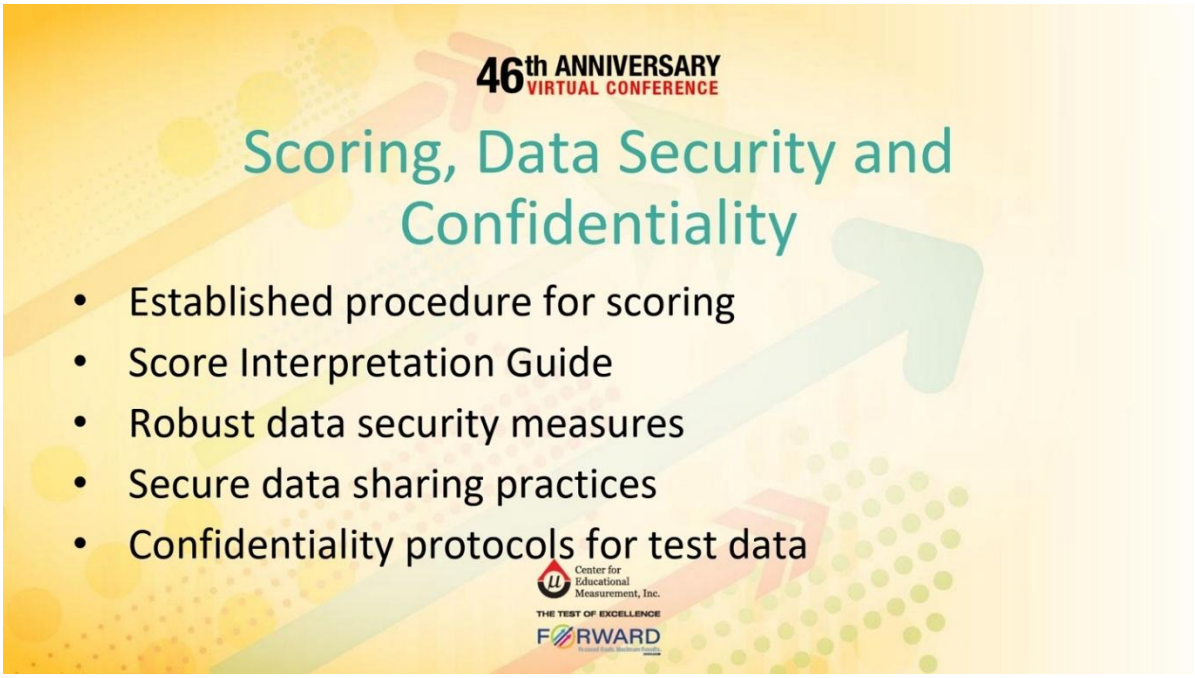
46th ANNIVERSARY VIRTUAL CONFERENCE

Comfortable Testing Environment for Examinees

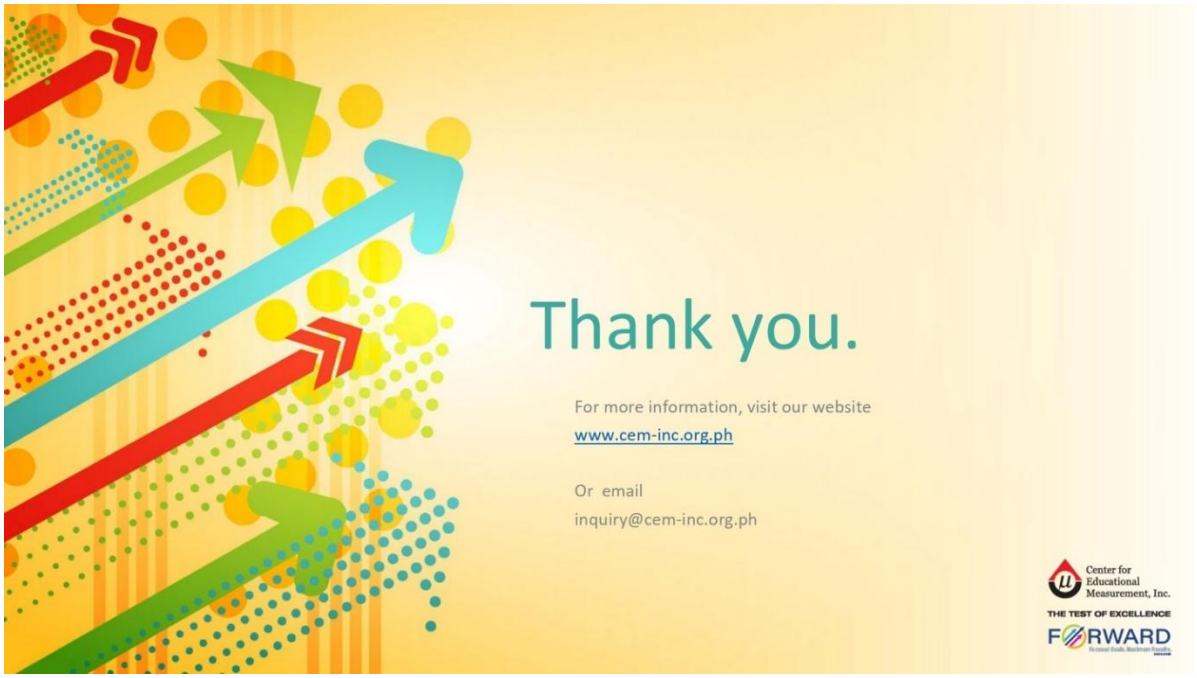
- Distraction-free setting
- Candidate orientation
- Practice opportunities
- Human proctor support
- Clear instructions

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

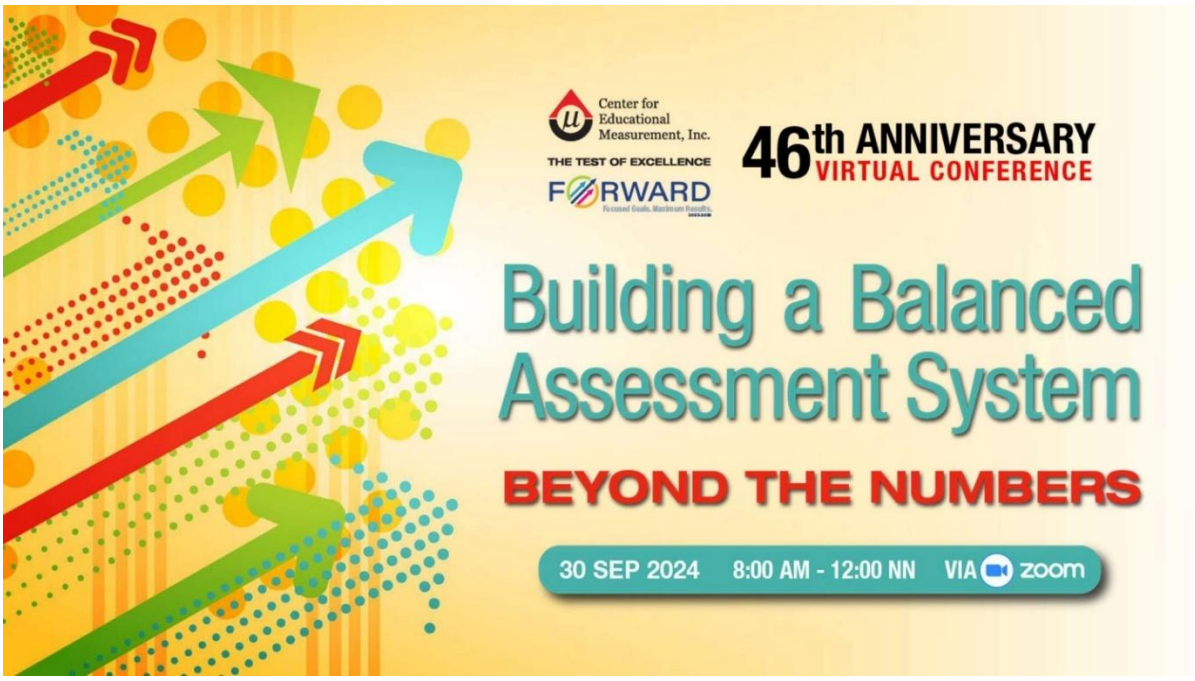

46th ANNIVERSARY
VIRTUAL CONFERENCE
Preventing Fraudulent Means of Obtaining Scores
• Random sampling design
• Continuous monitoring
• Test security
• Collaborative responsibility
Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

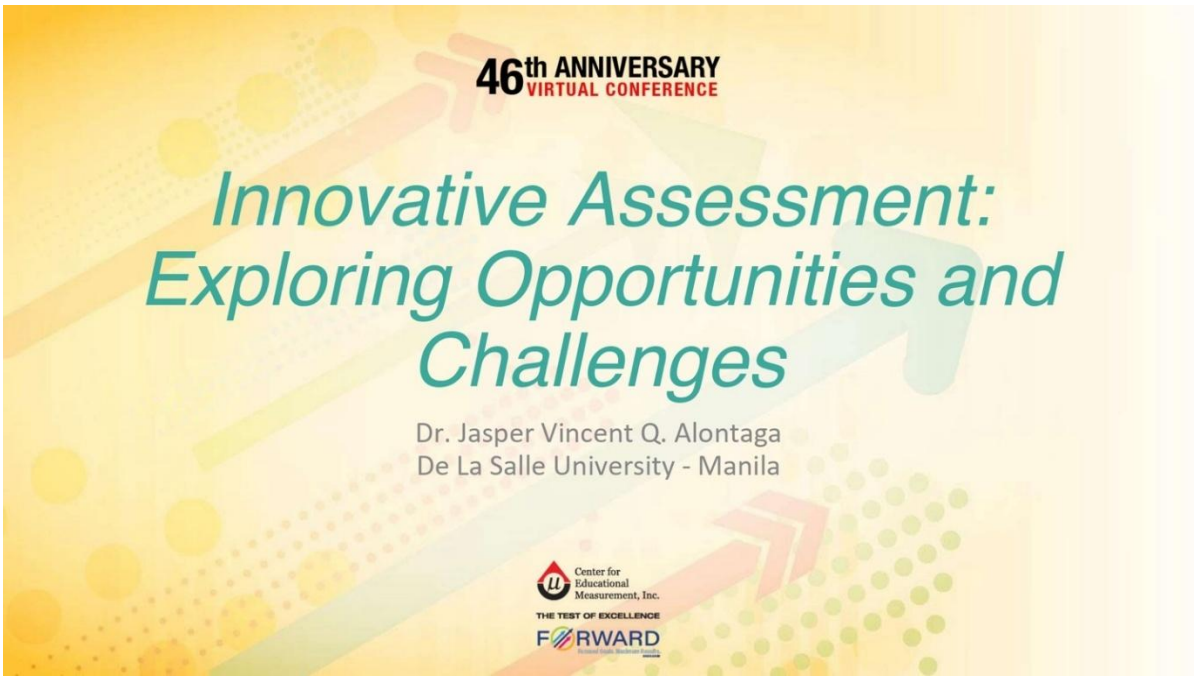

46th ANNIVERSARY
VIRTUAL CONFERENCE
Scoring, Data Security and Confidentiality
• Established procedure for scoring
• Score Interpretation Guide
• Robust data security measures
• Secure data sharing practices
• Confidentiality protocols for test data
Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

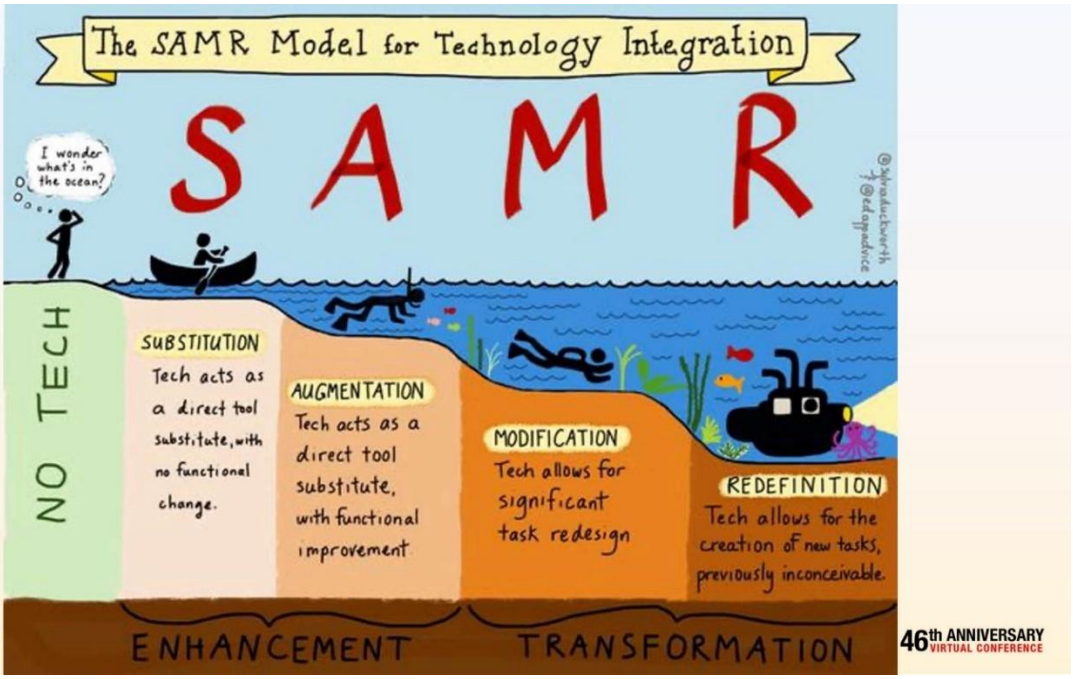

46th ANNIVERSARY
VIRTUAL CONFERENCE
The Standards
The International Association for Educational Assessment
International Standards for Educational Assessment Organisations
STANDARDS
for Educational and Psychological Testing
Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

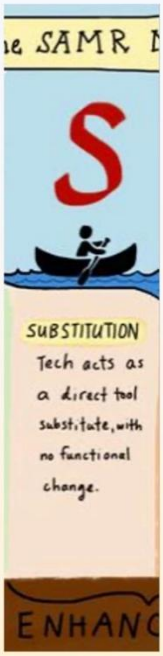

Thank you.
For more information, visit our website
www.cem-inc.org.ph
Or email
inquiry@cem-inc.org.ph
Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

Dr. Jasper Vincent Q. Alontaga









Online Exams

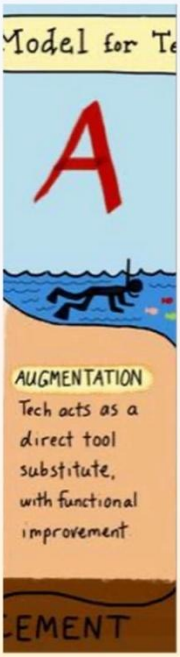
Managing and administering assessment can be more time efficient and accurate

Assessment results are available in a timely and actionable manner

Submission process becomes easier

Monitoring process can be done more frequently

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD
46th ANNIVERSARY VIRTUAL CONFERENCE



Variety

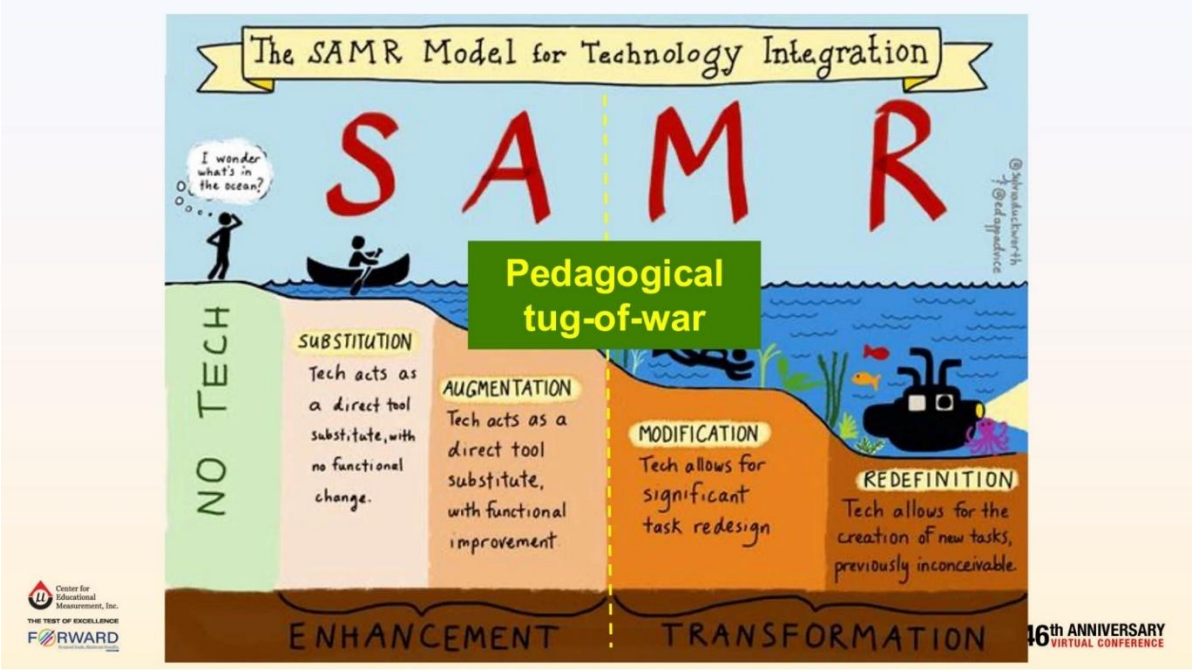
Gamify: Motivational assessment leads to mastery of competencies

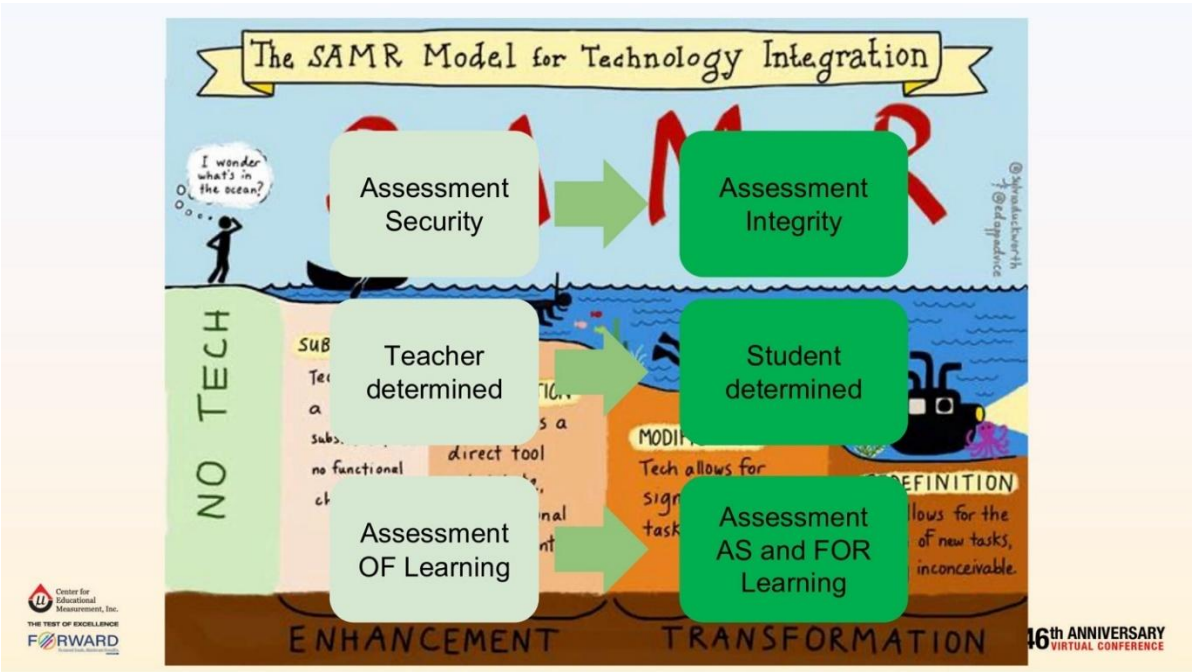
Multiple and flexible presentation, expression and engagement

K! FLIPQUIZ edpuzzle

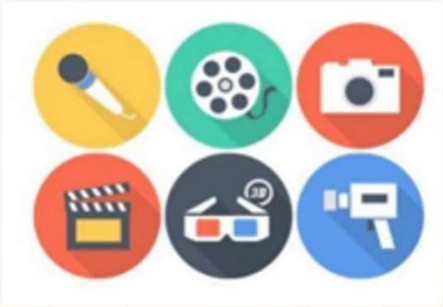
Considerations under S/A:

- ❑ Digital honor pledge
- ❑ Passcode and Time limits
- ❑ Test Banks. Randomizing items and choices
- ❑ AI as exam item generators
- ❑ Lock-down browser
- ❑ Online proctoring tools (live proctoring, remote AI-based red flags, audit logging)


The SAMR Model for Technology Integration
S A M R
I wonder what's in the ocean?
Pedagogical tug-of-war
NO TECH
SUBSTITUTION
Tech acts as a direct tool substitute, with no functional change.
AUGMENTATION
Tech acts as a direct tool substitute, with functional improvement
MODIFICATION
Tech allows for significant task redesign
REDEFINITION
Tech allows for the creation of new tasks, previously inconceivable.
ENHANCEMENT
TRANSFORMATION
Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD
46th ANNIVERSARY
VIRTUAL CONFERENCE


The SAMR Model for Technology Integration
I wonder what's in the ocean?
Assessment Security
Assessment Integrity
Teacher determined
Student determined
Assessment OF Learning
Assessment AS and FOR Learning
NO TECH
ENHANCEMENT
TRANSFORMATION
Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD
46th ANNIVERSARY
VIRTUAL CONFERENCE

Multimodal Digital Classroom Assessments



Teacher-designed assessment practices requiring students to combine two or more representational modes (media) using digital technology to create an **artifact** (product/performances)

Fjørtoft, H. (2020). Multimodal digital classroom assessments. *Computers & Education*, *152*, 103892.

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Product/Performance

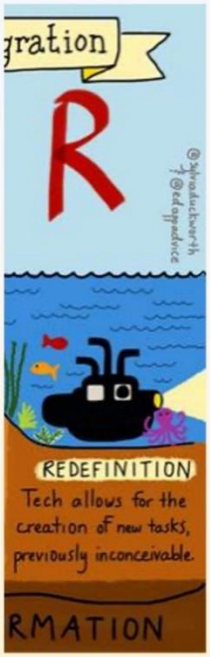
Demonstration of Learning

- Online Assignments
- Recorded or Live Performance of a Task
- Creation of a product
- Oral defense or explanation

Utilize tools to make assessment interactive and collaborative

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE



Project-based

Design more meaningful and authentic assessment task in ODL

E-portfolios: Digital artifacts maintained to exhibit abilities, achievement and growth

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

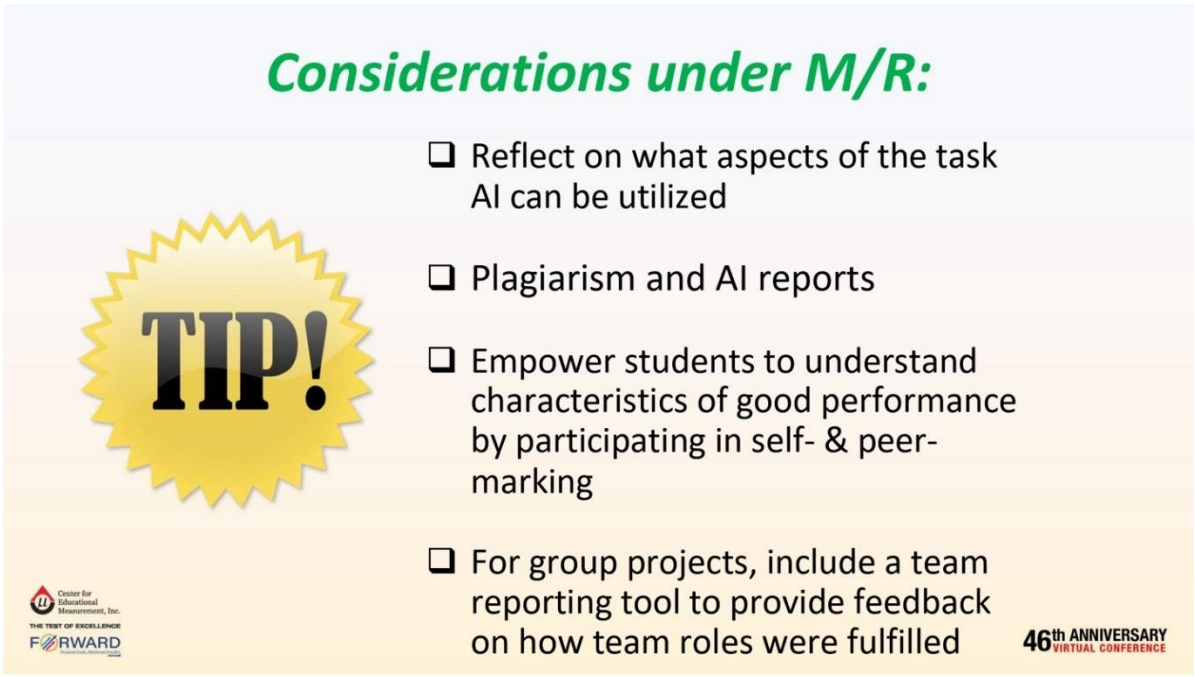
46th ANNIVERSARY VIRTUAL CONFERENCE

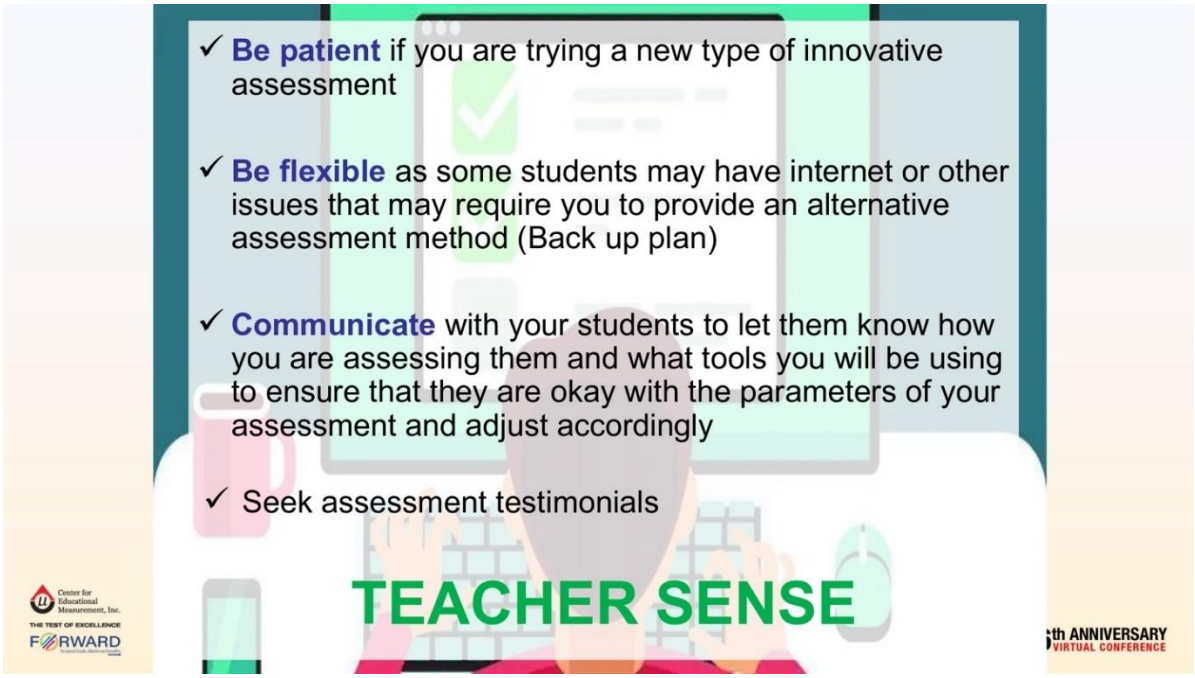
Negotiated Assessment

- ❑ Choice in groupings
- ❑ Choice in topics
- ❑ Choice in products
- ❑ Co-create criteria

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

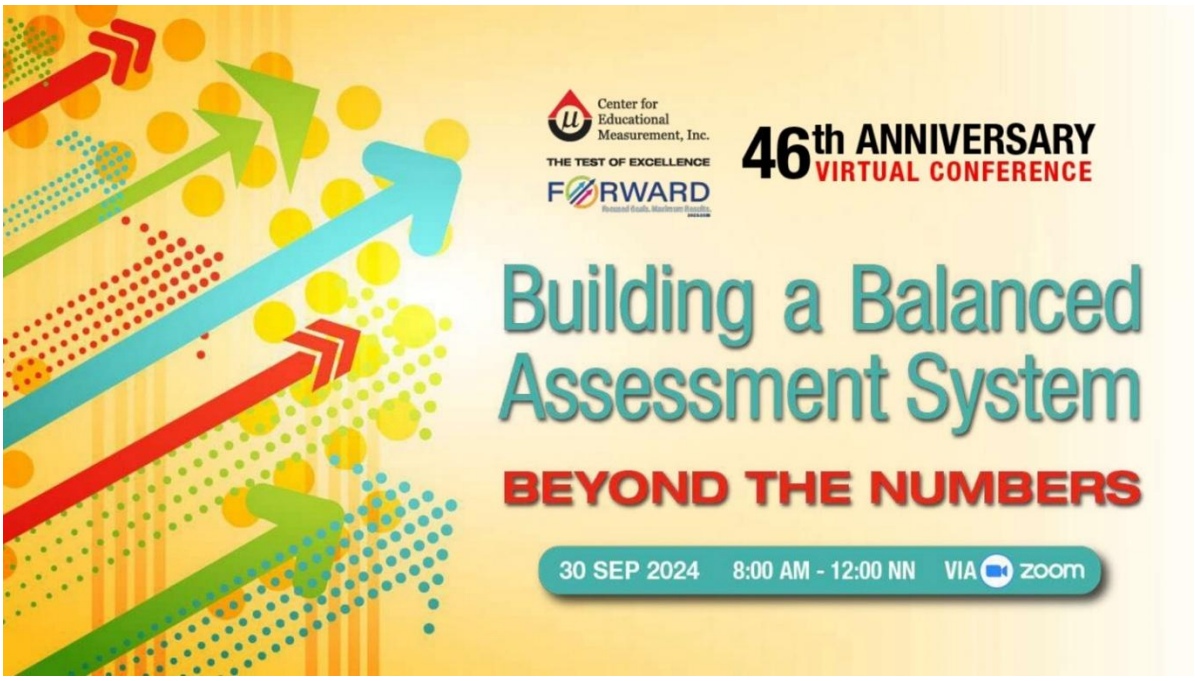
46th ANNIVERSARY VIRTUAL CONFERENCE

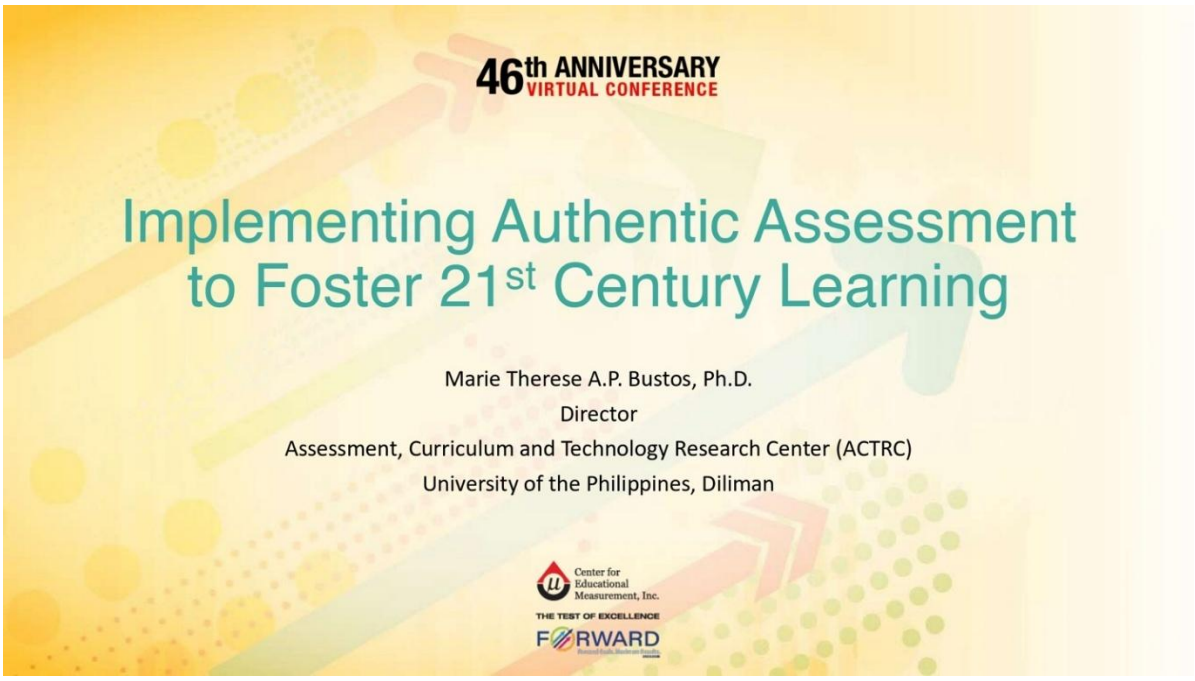

Considerations under M/R:
TIP!
❑ Reflect on what aspects of the task AI can be utilized
❑ Plagiarism and AI reports
❑ Empower students to understand characteristics of good performance by participating in self- & peer-marking
❑ For group projects, include a team reporting tool to provide feedback on how team roles were fulfilled
46th ANNIVERSARY VIRTUAL CONFERENCE


✓ Be patient if you are trying a new type of innovative assessment
✓ Be flexible as some students may have internet or other issues that may require you to provide an alternative assessment method (Back up plan)
✓ Communicate with your students to let them know how you are assessing them and what tools you will be using to ensure that they are okay with the parameters of your assessment and adjust accordingly
✓ Seek assessment testimonials
TEACHER SENSE


DE LA SALLE UNIVERSITY
RELIGIO MORES CULTURA
MANILA
Br. Andrew Gonzalez FSC
College of Education
ODEO
Online
Distance
Education
Office
Jasper Vincent Q. Alontaga, PhD
jasper.alontaga@dlsu.edu.ph
Join us:
Master of Arts in Education major in
Educational Technology
Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD
46th ANNIVERSARY
VIRTUAL CONFERENCE

Dr. Marie Therese A.P. Bustos





46th ANNIVERSARY VIRTUAL CONFERENCE

How should student learning outcomes be measured in SHS?

46th ANNIVERSARY VIRTUAL CONFERENCE

Standardized Tests

National Certifications

The proof of the pudding is in the eating.

Written tests and performance measures

Performance-based measures

Center for Educational Measurement, Inc.

THE TEST OF EXCELLENCE

FORWARD

46th ANNIVERSARY
VIRTUAL CONFERENCE

HIGHLIGHTS OF THE STUDY

1. **About 56% of TEIs in BLEPT elementary and secondary have passing rates below the 12-year average** national passing rate for overall takers.
2. First time takers have higher passing rates than repeaters. **Repeaters have higher likelihood of failing again.**
3. **Over 81% of COEs and 91% and CODs are not high-performing** or have overall passing rates below 75% in BLEPT elementary and BLEPT secondary.
4. Only **2% of TEIs are classified as high-performing** considering passing rates of overall takers.
5. Mindanao, **esp. BARMM** has the lowest passing rates in the country.
6. In most years, **Private Sectarian TEIs tend to outperform others** both in Elementary and Secondary

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY
VIRTUAL CONFERENCE

CalTPA Seminar

CalTPA
California Teaching Performance Assessment

Instructional Cycle 1: Learning About Students and Planning Instruction

Cycle Step	What You Need to Do	Evidence to Be Submitted
Step 1: Plan (templates provided, including a sample lesson plan format)	• Gather contextual information, including students' assets and learning needs, for the whole class and for each of the 3 focus students. • Develop one lesson plan using content-specific pedagogy, knowledge of your students, California Content Standards and/or Curriculum Frameworks, and California English Language Development Standards.* • Explain how the lesson plan addresses the learning needs of your students and is informed by relevant student assets, as well as how it uses UDL strategies to provide for an inclusive learning environment. • **Provide at least one ELD goal for English learners in your class—including those who are Fluent English Proficient (reclassified English learners), Heritage language speakers, or multilingual learners—and Focus Student 1 (FS1).** • Provide key instructional resources and/or materials related to the lesson plan (e.g., reading materials, graphic organizers, slides, support documents, educational technology).	• **Part A:** Written Narrative: Getting to Know Your Students (no more than 9 pages) • **Part B:** Lesson Plan (include content-specific learning goal[s] and ELD goal[s]) (no more than 10 pages) • **Part C:** Written Narrative: Lesson Plan Rationale (no more than 7 pages) • **Part D:** Related Instructional Resources and Materials (no more than 8 pages)
Step 2: Teach and Assess (annotation tool provided in the ePortfolio system)	• Teach the planned content-specific lesson to your students within the school placement. • Video record the full lesson. Select and annotate 3 video clips that show (1) how you create a positive and safe learning environment and establish expectations for content-specific learning; (2) how you engage students in activities and your instructional strategies; and (3) how you monitor students' learning.	• **Part E:** 3 Annotated Video Clips (no more than 5 minutes each)

Teacher evaluation: Marion & Buckley (2016)

- Use of Student Learning Objectives (SLO) as an accountability tool to supplement information coming from standardized approaches
- Parts of SLO
  - Learning goal
  - Instructional strategies to achieve the students' achievement of the goal
  - Targets for student and teacher performance
  - Assessments to be used to evaluate the learning goals

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

46th ANNIVERSARY VIRTUAL CONFERENCE

Authentic assessment

Performance – based assessment

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

- Authentic assessments are performance assessments using real-world or authentic tasks or contexts (Mueller, 2018)

46th ANNIVERSARY VIRTUAL CONFERENCE

- Authentic assessment values the student in the teaching and learning process.

46th ANNIVERSARY VIRTUAL CONFERENCE

Characteristics of Authentic Assessment (Wiggins, 1989)

- Structure and Logistics
- Intellectual Design Features
- Grading and Scoring

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Structure and Logistics

- Are more appropriately public; involve an audience, a panel and so on
- Do not rely on unrealistic and arbitrary time constraints
- Are more like portfolios (not one-shot)
- Require some collaboration with others
- Make assessment and feedback to students central

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Intellectual Design Features

- Are essential (not needlessly intrusive, arbitrary or contrived to "shake out a grade)
- Are enabling – constructed to point the student towards more sophisticated use of the skills and knowledge
- Are contextualized, complex intellectual challenges
- Are representative challenges designed to emphasize depth than breadth
- Involve somewhat ambiguous "ill-structured" tasks or problems

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

- Most of the important problems one faces are ill-structured, as are all the really important social, political, and scientific problems. . . . But ill-structured problems are not found in standardized achievement tests. . . . Efficient tests tend to drive out less efficient tests, leaving many important abilities untested and untaught. . . . All this reveals a problem when we consider the influence of an accountability system in education. . . . We need a much broader conception of what a test is. (Frederiksen 1984: 199)

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Grading and Scoring Standards

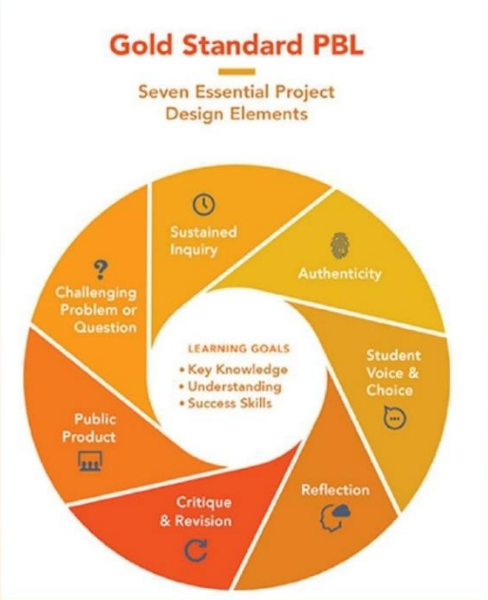
- Involve criteria that assess the essentials
- Graded in reference to performance standards and not the "curve"
- Use a multifaceted scoring system instead of one aggregate grade

- Multiple judges, when properly trained to assess actual student performance using agreed-upon criteria, display a high degree of interrater reliability.

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Authentic Assessment



Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Dimensions of authentic assessment
(adaptation of Mueller's work by UNSW)

DIMENSION	More traditional assessments	More authentic assessments
Structure of problems	Predetermined	Unpredictable
Learning setting	Contrived	Real
Cognitive activity	Lower-order	Higher-order
Learner agency	Teacher-defined	Learner-defined
Application of learning	Indirect evidence	Direct evidence

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

46th ANNIVERSARY VIRTUAL CONFERENCE

Expectation of a task
(Cumming and Maxwell, 1999)

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

First order expectation

- The development of knowledge and skills necessary to accomplish the task
- Facts, figures, concepts, principles, and "information" of the subject

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

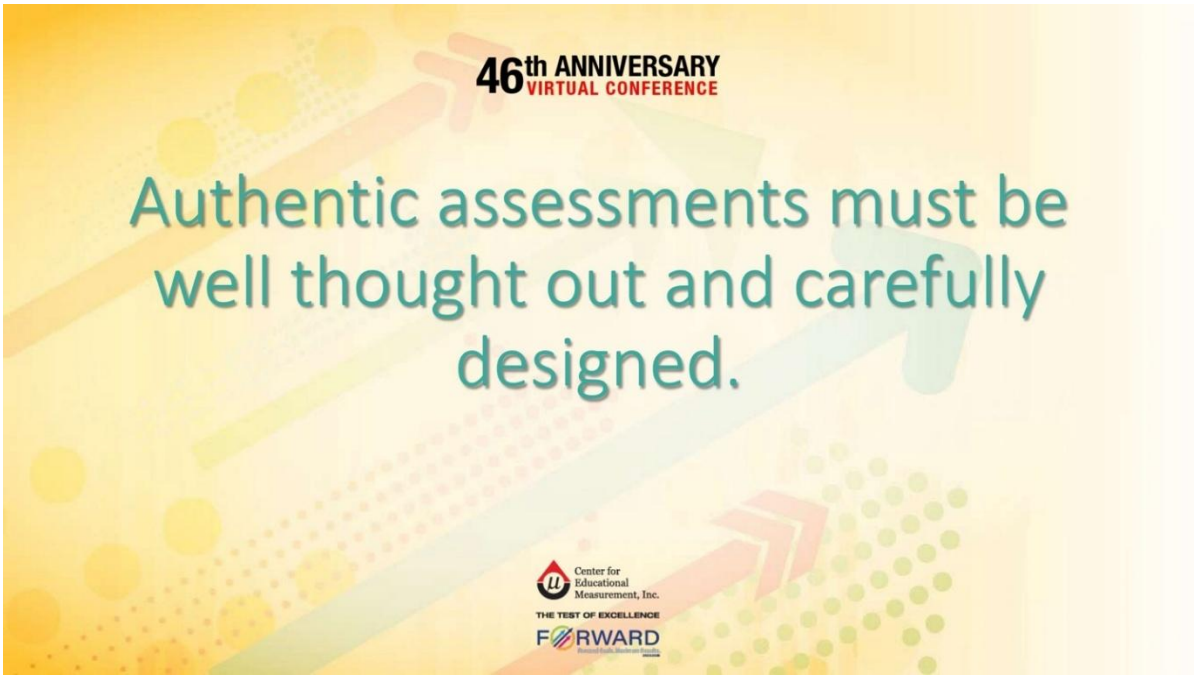
46th ANNIVERSARY VIRTUAL CONFERENCE

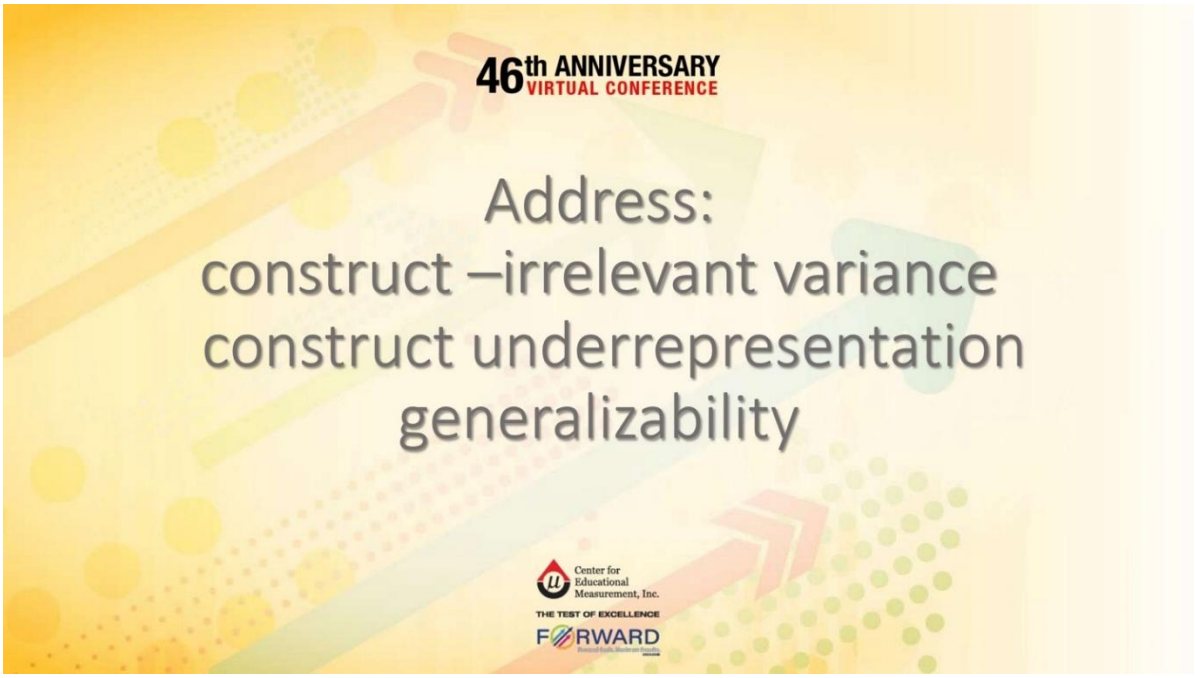
Second-order expectation

- The immersion in the activity and performance of actual behaviors relevant to the situation and subject area
- Application, evaluation, and synthesis are practiced and perfected

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE


46th ANNIVERSARY
VIRTUAL CONFERENCE
Authentic assessments must be well thought out and carefully designed.
Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD


46th ANNIVERSARY
VIRTUAL CONFERENCE
Address:
construct –irrelevant variance
construct underrepresentation
generalizability
Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

Reminders

- Carefully analyze the construct (relevant and irrelevant aspects of the task).
  - Specify the knowledge and skills required for the successful attainment of the learning goal.
- Make sure that the task and responses to the task fully represent the facets of the learning goal
- Enhance generalizability through tightly defined tasks and scoring rubrics and better training of raters
- Can't just have 1 task
  - Shavelston (1992): 6-12 performance tasks required to get a stable estimate of the student performance

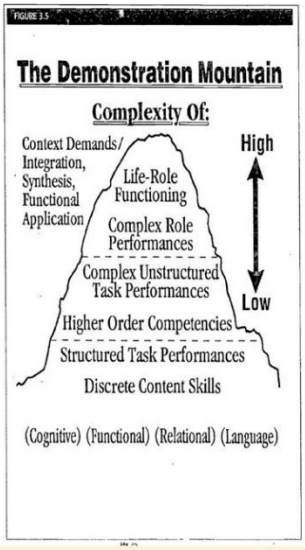
Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Demonstration Mountain of Performance (Spady, 1994)

Transformative

Traditional



Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Implications on assessment

- Assess with the long-term significant outcomes of the program in mind, not just subject-specific outcomes.

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

- Think photo albums instead of snapshots.

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

How to create authentic assessment (Mueller, 2018)

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Standard

- What should students know and be able to do?
- Based on the K to 12 curriculum / MATATAG

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Authentic Tasks

- What indicates students have met these standards?
- Brainstorm with a colleagues

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Criteria

- What does good performance on this task look like?
- Think evidence

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

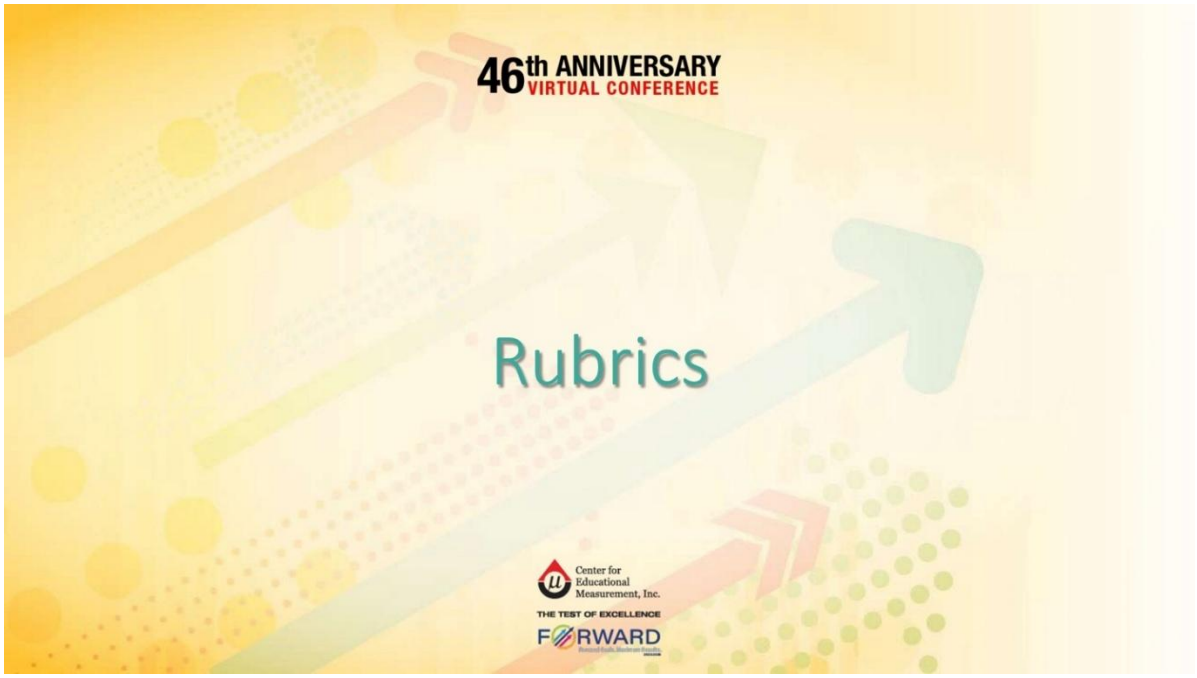
46th ANNIVERSARY VIRTUAL CONFERENCE

Rubric

- How well did the students perform?
- Criteria and the levels of performance

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

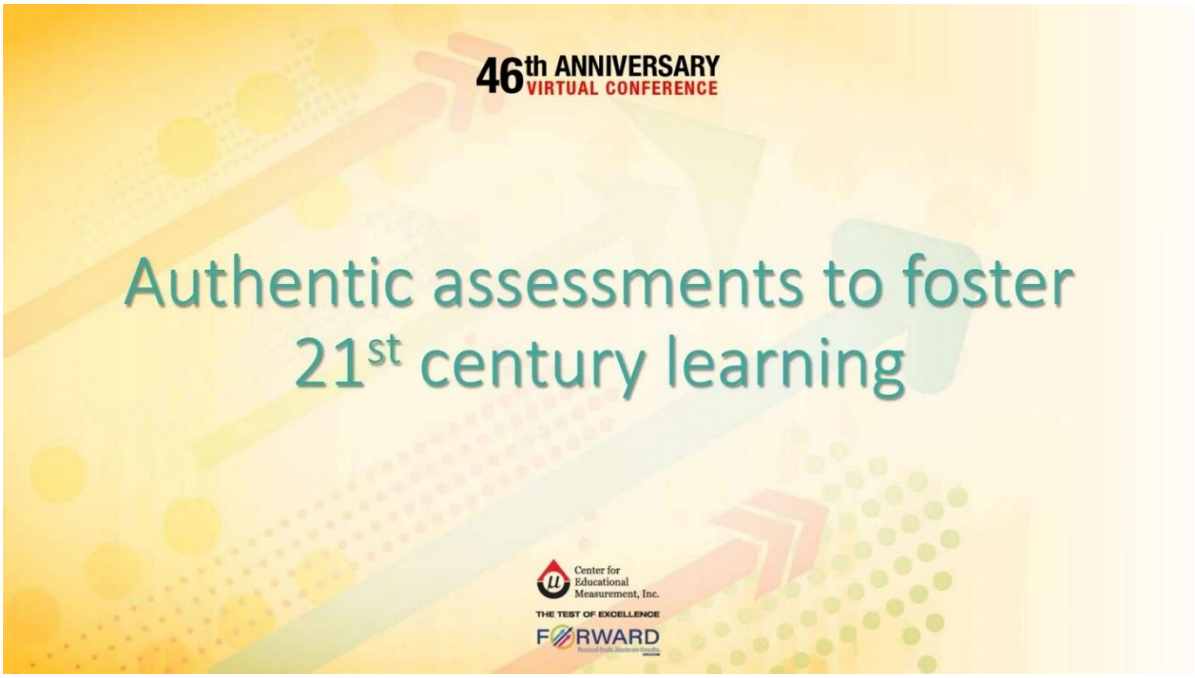


Guidelines for developing rubrics (HK PolyU, 2018)

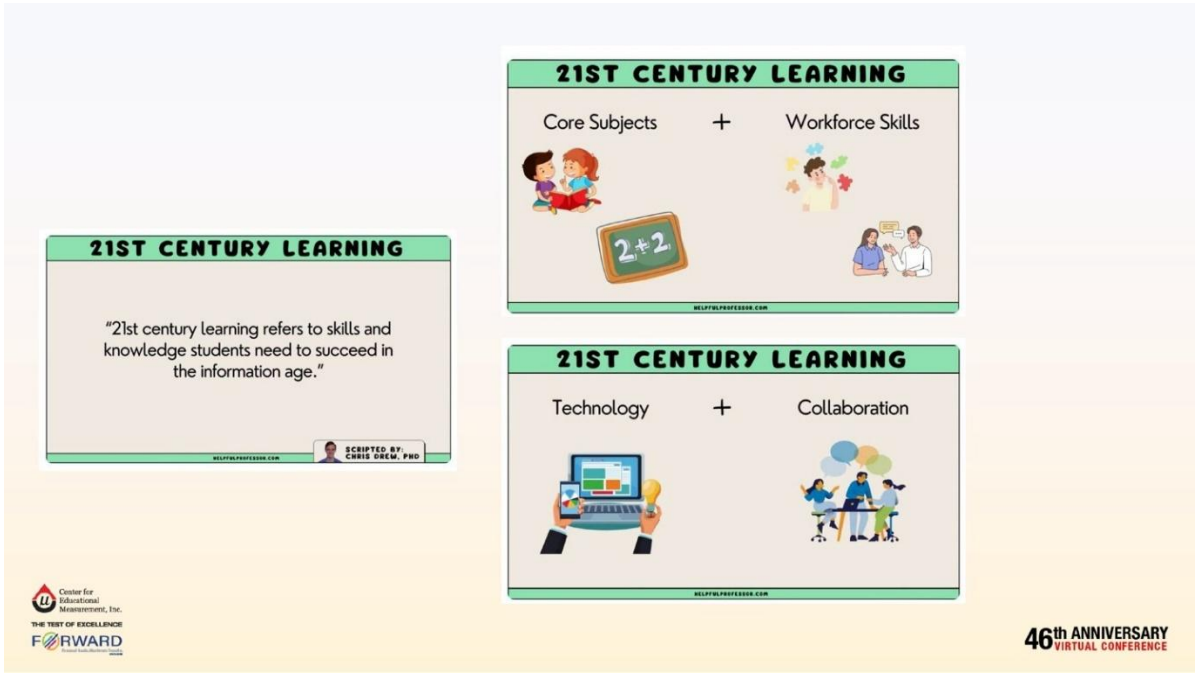
- Step 1 - Identify the purpose and aims of assessing students
- Step 2 - Identify what to assess
- Step 3 - Select an appropriate type of rubric
- Step 4 - Identify the performance criteria for assessing student work
- Step 5 - Identify the levels of performance
- Step 6 - Describe each level of performance (grading descriptors)
- Step 7 - Pilot the rubrics
- Step 8 - Periodical review / revisions to rubrics as necessary

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

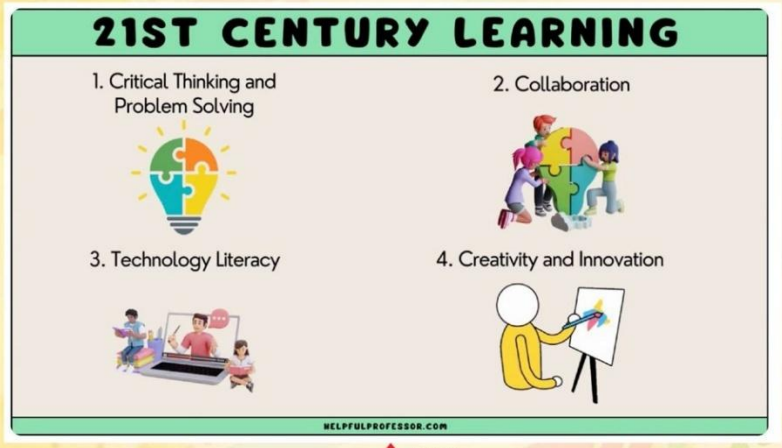
46th ANNIVERSARY VIRTUAL CONFERENCE




Information, Media and Technology Skills
Learning and Innovation
Communication Skills
Life and Career Skills
DepEd's 21st Century Skills
Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD


21ST CENTURY LEARNING
"21st century learning refers to skills and knowledge students need to succeed in the information age."
SCRIPTED BY: CHRIS DREW, PHD
21ST CENTURY LEARNING
Core Subjects + Workforce Skills
2+2
21ST CENTURY LEARNING
Technology + Collaboration
Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD
46th ANNIVERSARY
VIRTUAL CONFERENCE

46th ANNIVERSARY VIRTUAL CONFERENCE



Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Situational Judgment Tests

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

FIGURE 1.
Sample SJT Item

You and a colleague from another department are jointly responsible for coordinating a project involving both departments. Your colleague is not completing an appropriate portion of the work. What should you do?

A. Ask your supervisor to discuss the problem with your colleague's supervisor.
B. Remind your colleague that the project will not be completed effectively without effort from both of you.
C. Tell your colleague that you will discuss the problem with your colleague's supervisor if your colleague refuses to work on the project.
D. Tell your colleague that nonparticipation creates more work for you and makes it harder to finish the project.
E. Ask someone else from your colleague's department to help with the project.

(Whetzel; Sullivan; and McCloy, 2020)

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

Situational Judgment Tests

- Situational judgment tests (SJTs) assess individual judgment by presenting examinees with problem scenarios and a list of plausible response options.
- SJTs tap into various relevant constructs through the realistic scenarios presented and the accompanying questions.

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Constructs that can be measured through SJTs

- Communication
- Teamwork and collaboration
- Leadership and influence
- Compassion
- Empathy
- Conflict management
- Problem solving
- Negotiation
- Resilience
- Ethical responsibility
- Self-awareness
- Adaptability
- Reliability

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Reminders

- Use critical incidents as scenarios.
- Be careful with transparency of options.
- Address construct validity.

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

46th ANNIVERSARY VIRTUAL CONFERENCE

Situational Judgment Tests and SOLO Taxonomy

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

Structure of Observed Learning Outcomes SOLO Taxonomy

Table 1.1: SOLO levels, symbols and hand signs

Prestructural	Unistructural	Multistructural	Relational	Extended abstract
				
Learning outcomes show unconnected information and no organisation.	Learning outcomes show simple connections but importance is not noted.	Learning outcomes show connections are made but significance to overall meaning is missing.	Learning outcomes show connections are made and parts are synthesised with the overall meaning.	Learning outcomes go beyond the subject and make links to other concepts – generalising, predicting, evaluating.
No idea	One idea	Many ideas	Related ideas	Extended ideas
				

https://www.teachingtimes.com/building-students-reflective-skills-with-solo-taxonomy/

46th ANNIVERSARY VIRTUAL CONFERENCE

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

QS4_Item1

Evaluation 1: The teacher gave a short recap of the topic that was discussed the previous day. As she starts asking questions about it, no student is attempting to answer. What do you think should she do?

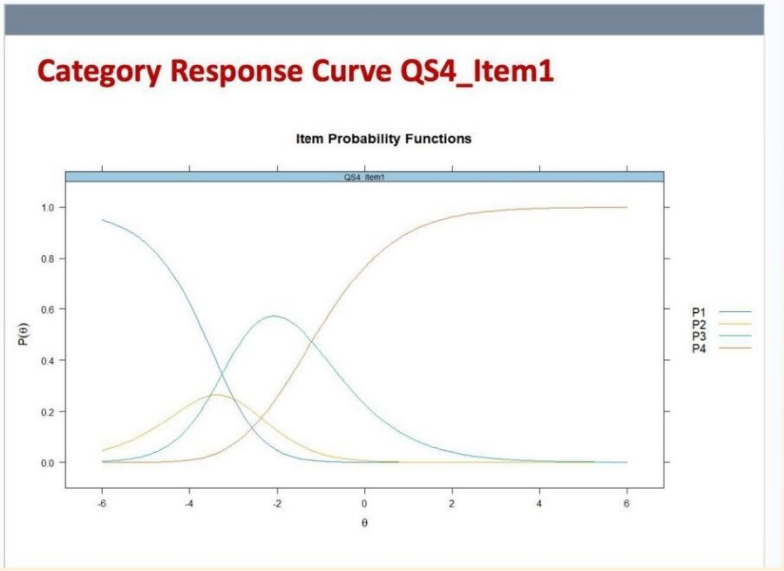
OPTIONS	RATIONALE
A. The students did not learn the lesson, so she should teach the topic again if there's still time.	1 – This action focuses on the typical lesson delivery.
B. The students are just shy, so she should just proceed with her lesson plan.	0 – This action views student data as personal characteristics and proceeds with usual lesson delivery.
C. The teacher should change her question; consequently, she might get a response from the students.	2 – This action tries to elicit additional student data but with no clear direction on how to understand the learning gaps.
D. The teacher should ask questions with different levels of complexity to know where the students are.	**3 – This action tries to vary the eliciting prompts to know the current level of students in order for the teacher to understand the learning gaps.**

(ACTRC, 2021)

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Category Response Curve QS4_Item1



Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

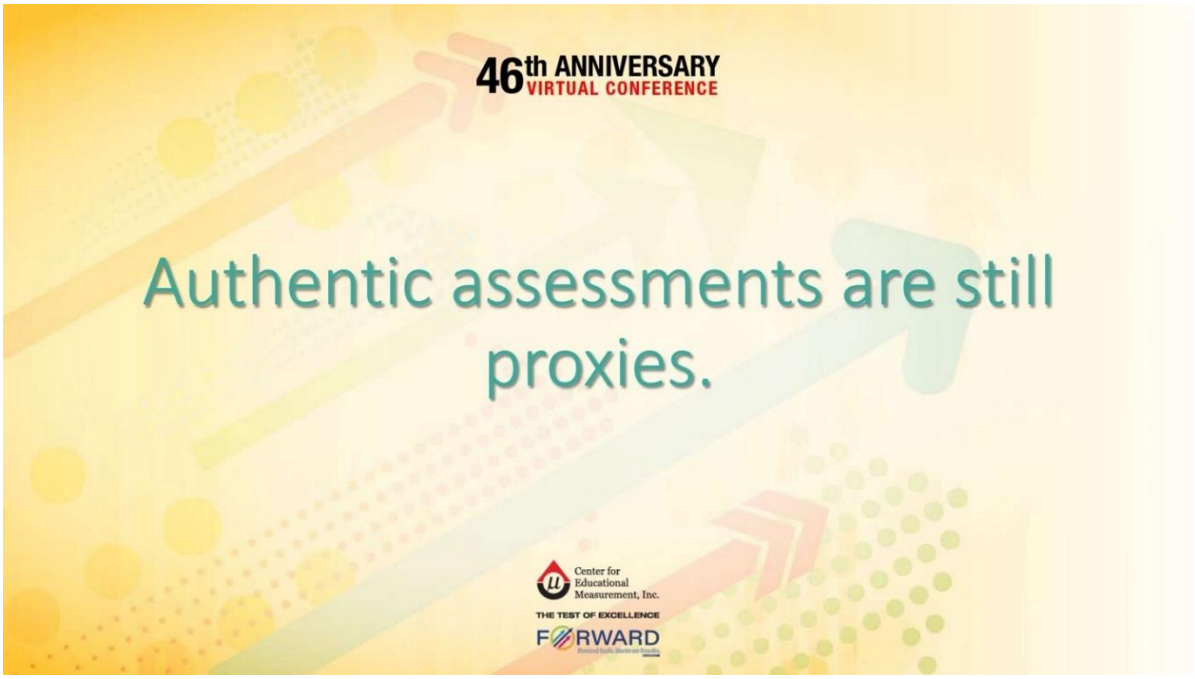
46th ANNIVERSARY VIRTUAL CONFERENCE

Cagasan and Santos (2024)

- The hypothesized categories in SOLO items need to be empirically validated to see if these are functioning.
- Caution is recommended in using it for grading purposes especially if the SOLO items are not empirically validated.

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE



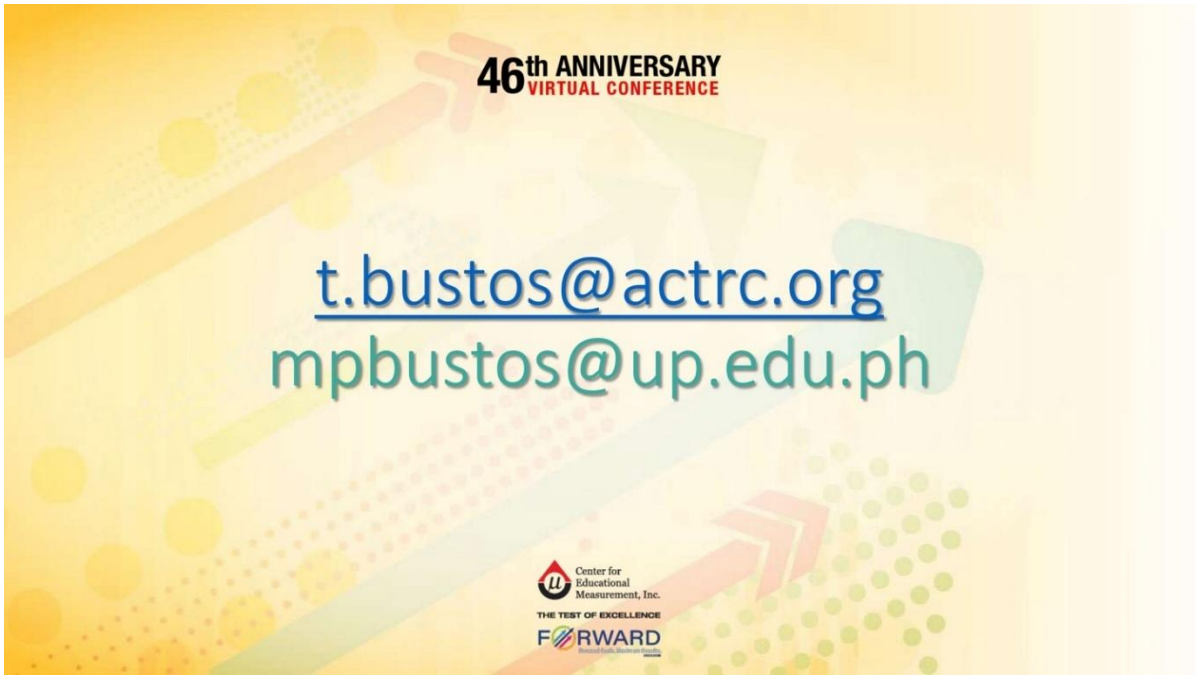
However, performance-based (authentic) assessments:

- May be the only way to measure some intended constructs
- May be the better way to measure some intended constructs
- Produce instructional information in addition to accountability information
- Provide learning and assessment opportunities for students
- Signal the types of instructional tasks we want to see in classrooms.

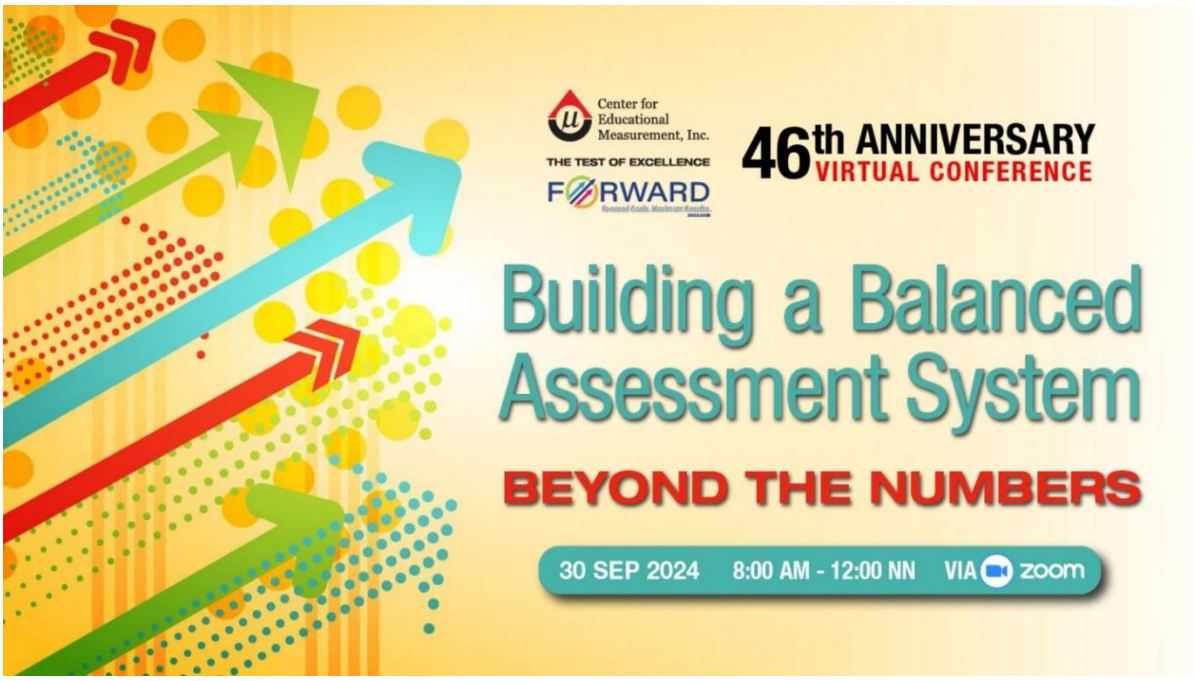
Marion & Buckley (2016)

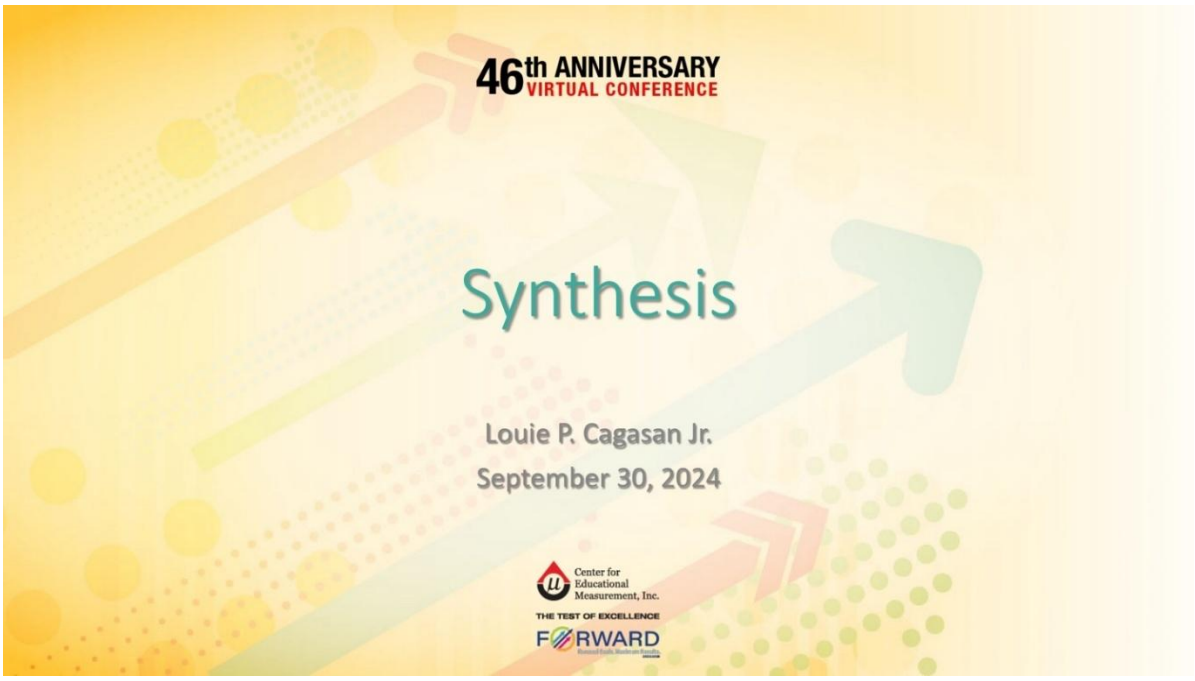
Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE


46th ANNIVERSARY
VIRTUAL CONFERENCE
t.bustos@actrc.org
mpbustos@up.edu.ph
Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

Mr. Louie P. Cagasan, Jr., MAPsych





Balanced Assessment System

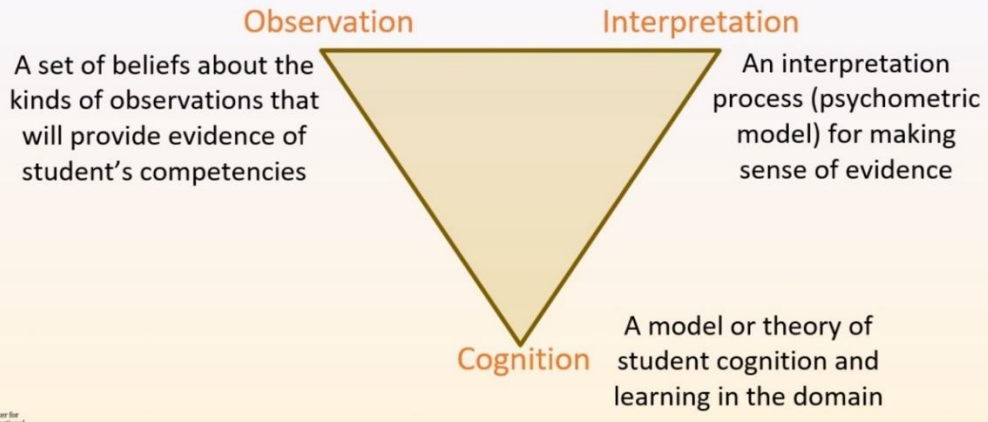
- Comprehensive
- Continuous
- Efficient
- Useful
- Coherent (Vertically and Horizontally)

Dela Torre (2024)

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Designing and Implementing a Balanced Assessment System: The Assessment Triangle



The Assessment Triangle

Dela Torre (2024)

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Interdependence of Components

There's a growing belief among many educational assessment experts that if assessment, curriculum and instruction were more integrally connected student learning would improve.

(Pellegrino, 2017, p. 364)

Source: Griffin, Cagasan, Care, Vista, Nava, 2016, p 76

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Assessment can act as Leverage for Quality Education

- Informs planning and educational decisions
- Enhances that process of education by providing feedback that supports teaching and learning
- Substantiates the outputs and outcomes of education

Tan (2024)

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Standardized tests with established reliability and validity evidence are essential to have trustworthy data that can be used for actionable plan/s.

CEM can provide these educational assessment needs.

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

SAMR Model for Technology Integration

- Substitution
- Augmentation

} Enhancement

- Modification
- Redefinition

} Transformation

There's a natural progression from no technology to transformation.

Alontaga (2024)

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

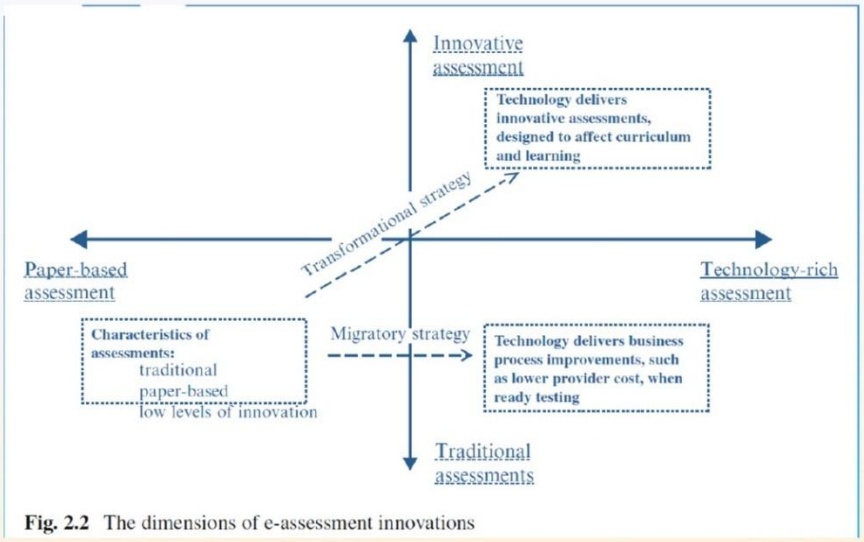


Fig. 2.2 The dimensions of e-assessment innovations

Binkley et al. (2012) pp. 27

Authentic Assessment

- Authentic assessments must be well thought out and carefully designed.
  - need to pilot our assessments (rubric, tasks, and innovations).
  - need to establish validity
- Authentic assessments are still proxies.

Bustos (2024)

Authentic Assessment

Authenticity lies in an assessment requiring students to use the same competencies, or combinations of knowledge, skills and attitudes, that they need to apply in the criterion situation in professional life. The level of authenticity of an assessment is thus defined by its degree of resemblance to the criterion situation.

(Gulikers et al., 2004, p. 69 as cited by Care and Kim (2018).

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Assessment of 21st Century Skills

In comparison to the educational assessment of content based knowledge, assessment of twenty-first century is still in its **infancy**… **Challenges** in assessing twenty-first century skills lie in **our lack of comprehensive understanding** of the nature and development of skills, about their multidimensionality, and about how to partition variance in behavior that is attributable to knowledge , or attributable to skill (Care & Kim, 2018, p. 23)

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

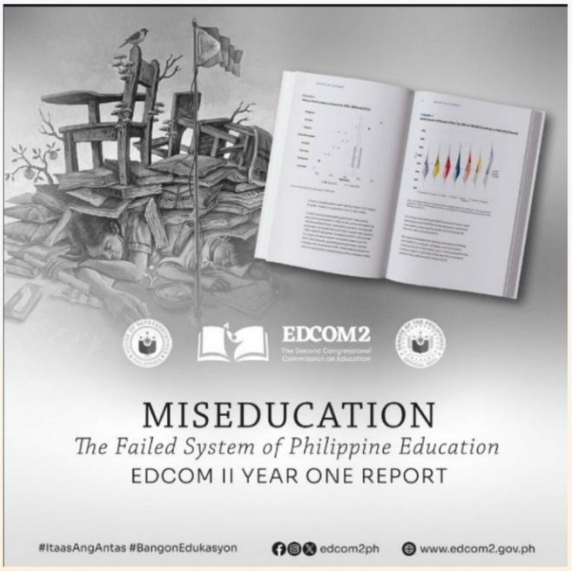
Assessment of 21st Century Skills

- How malleable are skills? (Soto, 2021; Soto 2022; Pancorbo,…, John…, De Fruyt, 2021)
- How to best measure these skills? (Duckworth & Yeager, 2015; Soto, 2021; Soto 2022; Pancorbo,…, John…, De Fruyt, 2021)
- How are these skills related to other constructs? (Soto, 2021; Soto 2022)

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

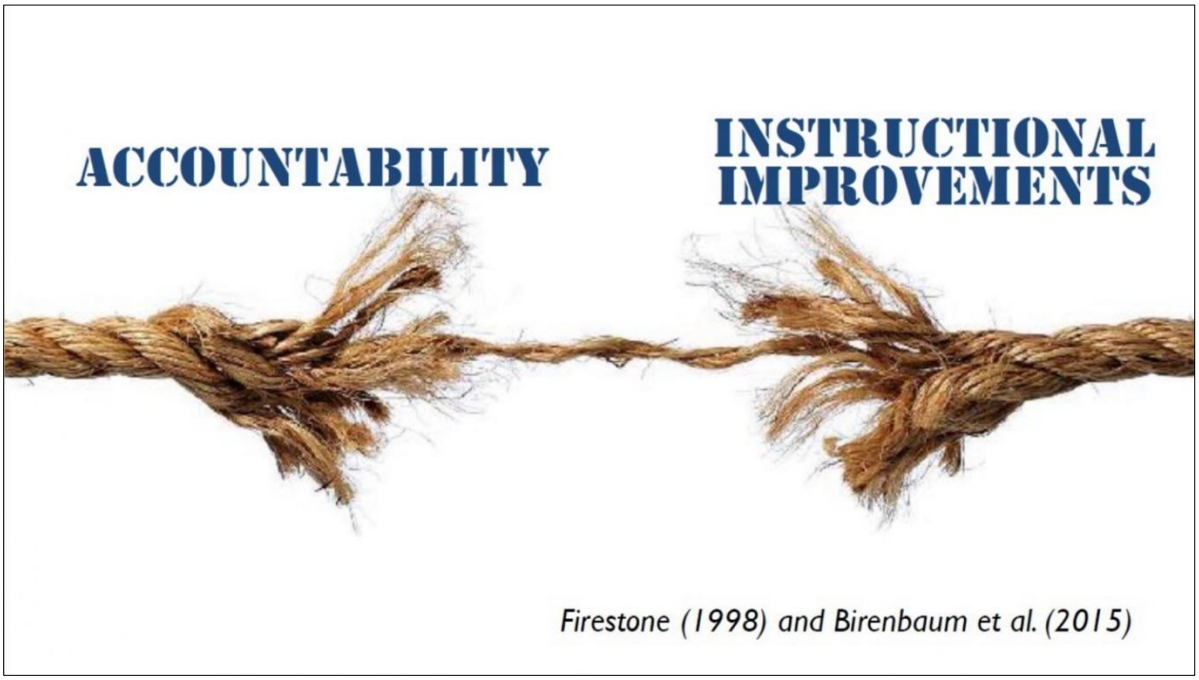
Where are we right now?

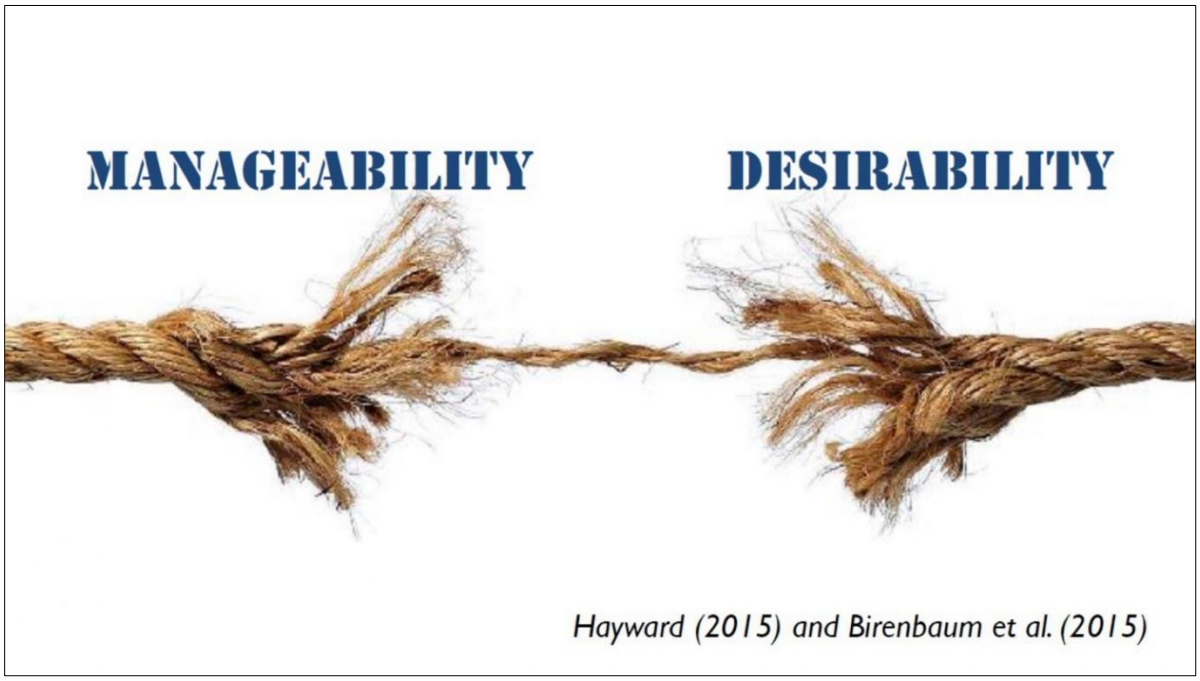


Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE


2 Common Tensions in Assessment
Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD


ACCOUNTABILITY
INSTRUCTIONAL IMPROVEMENTS
Firestone (1998) and Birenbaum et al. (2015)



How to get there?

I. Address the tensions, weaken the blocking forces, and strengthen the facilitative forces.

II. Be acquainted and use advanced and robust concepts and tools.

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Reminder: The Need for Space

Laveault (2016) believes that there should be space for policy developers and policy implementers to "co-regulate" or make appropriate adjustments and accommodation

Co-regulation between the policy (or policy makers) and practitioners (or teachers)

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Insanity: doing the same thing over and over again and expecting different results.

Albert Einstein

"When you fail to plan, you are planning to fail"

– Benjamin Franklin

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

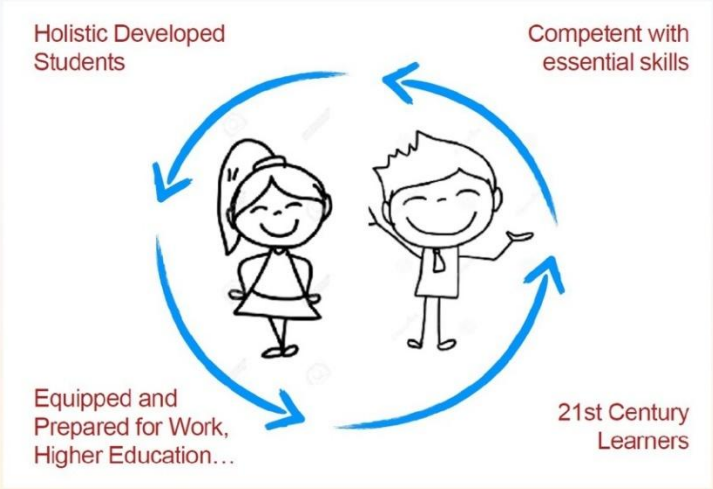
46th ANNIVERSARY VIRTUAL CONFERENCE

When there is no vision,
people perish

Proverbs 29:18a

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE



Para sa Bata, Para sa Bayan

Center for Educational Measurement, Inc.
THE TEST OF EXCELLENCE
FORWARD

46th ANNIVERSARY VIRTUAL CONFERENCE

Center for Educational Measurement, Inc.

THE TEST OF EXCELLENCE

FORWARD

Focused Goals. Maximum Results.

www.cem-inc.org.ph