rEFLections
Vol 31, No 2, May - August 2024

Input as a Key Element in Test Design: A Narrative of Designing an Innovative Critical Thinking Assessment

KHAGENDRA RAJ DHAKAL*
School of Liberal Arts, King Mongkut's University of Technology Thonburi, Thailand
RICHARD WATSON TODD
School of Liberal Arts, King Mongkut's University of Technology Thonburi, Thailand
NATJIREE JATURAPITAKKUL
School of Liberal Arts, King Mongkut's University of Technology Thonburi, Thailand
Corresponding author email: raj.kmutt@gmail.com

Article information
Article history:
Received: 17 Nov 2023
Accepted: 12 Jan 2024
Available online: 7 Aug 2024

Keywords:
Soft skill
Test input
Item design
Test development
Innovative test
Critical thinking

Abstract
Test input has often been taken as a given in test design practice. Nearly all guides for test designers provide extensive coverage of how to design test items but pay little attention to test input. This paper presents the case that test input plays a crucial role in designing tests of soft skills that have rarely been assessed in existing tests. In the process of designing a test of critical thinking, several attempts following existing test design guides resulted in poor tests that did not truly assess the intended objectives. These initial attempts used the norm of short passages as test input. Following these failures, we switched to using real-world input, such as tweets, numerical tables, and spam emails. In doing so, it was found that a particular input type favored a particular sub-skill of critical thinking and a particular item type. For example, using tweets as input enabled the assessment of the Perspective Taking sub-skill of critical thinking. This paper concludes that in designing skill tests, integrating appropriate input is at least as important as item design and calls for reevaluating the functions of test input as a distinct and dynamic element.

INTRODUCTION

Despite a massive oeuvre of research on test design and effectiveness, the impact of test input in skill assessments remains inadequately addressed. Test input refers to "the material that accompan[ies] a set of items and that the examinee must use to answer the questions or problems posed by the item" (Brookhart & Nitko, 2014, p. 9). In this article, we are focusing on test tasks where a single test input is used as the source for answering several test items. This is a frequently used test format, especially in skills tests, and is perhaps most familiar in tests of reading comprehension. As the focus in learning science shifts from knowledge recall to skill assessments (Brookhart & Nitko, 2014; Doye, 1991; Lewkowicz, 2000), such test input as the prompt for numerous test items becomes more important, and we will argue that this test input is crucial for enhancing test authenticity.

Test authenticity can be viewed from two perspectives. Psychometric authenticity provides evidence of consistency in the scores and is related to reliability. Situational authenticity, the type of authenticity we focus on in this article, concerns the extent to which test tasks match real-world tasks and, thus, is crucial for ensuring that the test results accurately reflect test takers' real-world abilities and can be used to make valid and reliable decisions based on the test scores.

This paper narrates the journey of the first author in creating the Multi-Purpose Assessment of Critical Thinking (MPACT) test, an innovative critical thinking assessment. The first attempt to design a critical thinking test based on following the testing literature and existing tests was largely a failure. To redesign the test to overcome the problems of the first draft, the first author focused on test input, especially the authenticity of this input. With this focus on test input, the second draft was far more successful in achieving the test's objectives. In addition to narrating the process of test development, this article also aims to examine how test input can enhance the authenticity, validity, and fairness of the MPACT test.

The narrative of the test designer

1. The origin of the MPACT test

This section covers the origin of the MPACT test, charting its evolution from the initial spark of inspiration to a structured idea. It covers the rationale behind the inception of the MPACT test, explaining why and how it was initiated. The narrative also outlines the development of a knowledge base that was methodically assembled to guide the design of the test.

1.1 The background of the MPACT test design

The conception of the MPACT test design emerged during a workshop on soft skills facilitated by the second author of this paper. The need for a valid test to evaluate students' critical thinking was evident, particularly in the context of complex global workplace needs. Despite my decade-long teaching experience and familiarity with formative and summative assessments, the task of designing a critical thinking test was a novel challenge. Guided and supported by the second and third authors as commentators and evaluators, I dedicated five years to completing the iterative process of researching, designing, and piloting the MPACT test. The test was designed primarily as a proficiency test of critical thinking, and this main purpose guided the design process. Having completed and evaluated the final version of the test, however, it appeared that the test can also be used to serve placement and diagnostic purposes as secondary uses, hence the use of 'Multi-Purpose' in the name of the test. This paper shares these experiences in a narrative format to benefit the testing community.

1.2 Gaining a knowledge base

The process of designing the MPACT test began with cultivating a comprehensive understanding of test design. I reviewed the following five textbooks on test design for theoretical and practical insights.

1. *Measurement and Evaluation in Education and Psychology* by Mehrens and Lehmann (1991).
2. *Language Testing in Practice* by Bachman and Palmer (1996).
3. *Designing and Analyzing Language Tests* by Carr (2011).
4. *Developing and Validating Test Items* by Haladyna and Rodriguez (2013).
5. *Educational Testing and Measurement* by Kubiszyn and Borich (2013).

These resources, relevant to my professional involvement in applied linguistics and general education, provided detailed guidance for test development. They agreed on the need to identify broad goals contextualized to the test's use and purpose, define specific objectives, develop test items, and evaluate the test.

2. The initial design of the MPACT test

The following sections detail the experiences encountered in the initial design process of the MPACT test, framed within these four steps.

2.1 Identifying the broad goals for the MPACT test

The 21st century's shift towards a complex global economy with constant disruption has changed priorities in education and employment. Rather than preparing students for stable, predictable employment, educational institutions now focus on equipping students for an uncertain future and a changing landscape of labor economics, emphasizing soft skills, particularly in tertiary education (Heckman & Kautz, 2012; Hora, 2019; Majid et al., 2012; Rios et al., 2020). Critical thinking stands out as a vital soft skill across various contexts considered a core competency in tertiary education (OECD, 2018; Scully, 2017) and key to academic success (Conley, 2008; Gomez, 2002; Liu et al., 2014). More importantly, employers of university graduates recognize critical thinking as a top employability skill (Casner-Lotto & Barrington, 2006; Ellerton & Kelly, 2022; Rios et al., 2020; Suarta et al., 2017). Furthermore, higher levels of critical thinking correlate with fewer negative life events (Butler, 2012), making it a significant objective in tertiary education and a focus of assessment.

2.1.1 The need for a new test of critical thinking in Thailand

Critical thinking, globally recognized as essential in diverse contexts, currently lacks an authentic test that gauges it in real-world scenarios. The necessity to combat challenges like disinformation, conspiracy theories (WEF, 2013) and online scams (Cross et al., 2016; PwC, 2022), to understand conflicting perspectives (Galinsky, 2010), and to use numerical data effectively (Jain & Rogers, 2019) underscores the importance of critical thinking skills. These skills are crucial for higher education students to effectively negotiate today's dynamic work environment.

Thailand's higher education sphere echoes the global requirement for an authentic test for critical thinking. The nation faces a mandate to bolster critical thinking in university graduates to foster a competitive workforce (ONEC, 2003; Pillay, 2002). Previous studies advocate prioritizing critical thinking in higher education for educational reforms and bolstering Thailand's global competitiveness (Buranapatana, 2006; Chaitrong, 2019; Pathanasethpong,

2014; Ploysangwal, 2018). However, Thailand lacks a clear definition of critical thinking skills and an integrated assessment framework for higher education. There is a void of locally designed critical thinking tests for purposes like hiring, professional growth, and promotion. To fill this gap, based on my extensive experience in Thai universities, I resolved to create the MPACT, specifically targeting college students in Thailand.

2.1.2 Purpose of the MPACT test: Authentic proficiency test of critical thinking.

The MPACT test's primary goal is to provide an authentic proficiency assessment of critical thinking, with particular emphasis on professional applicability. Envisaged as a multi-functional tool, the MPACT test is designed to appraise critical thinking competencies useful for tasks such as employee selection, placement, career development, and promotion across varied industries and organizational environments. It acknowledges the pivotal role of critical thinking in managing the intricacies of modern workplaces and adapting to the dynamic labor market's skill requirements. By addressing known deficiencies in current critical thinking assessments and by accurately mirroring tasks and contexts within contemporary professional settings, the MPACT test aspires to become a robust and authentic critical thinking measure. To guide the test design, I operationalized critical thinking (Dhakal, 2023) as:

> Critical thinking is an informed decision-making process for forming beliefs on a complex issue, solving real-world problems, and planning for probable futures by using an iterative approach that includes identifying biases, recognizing the quality of evidence and lapses in reason, spotting the flaws in quantitative data, pointing the logical fallacies, and evaluating other's perspectives contextually. (p. 21)

2.2 Defining specific objectives

This section focuses on defining the specific objectives of the MPACT test by identifying the most relevant and distinct sub-skills of critical thinking, which form the backbone of the test. The section details the exploration of these sub-skills and identifies the key sub-skills of critical thinking.

2.2.1 Defining specific sub-skills of the MPACT test

The main objective of designing the MPACT test was to create an authentic test of critical thinking that addresses the limitations of existing tests and encompasses the most relevant sub-skills for the 21st-century workforce.

A critical review of the major existing critical thinking tests was carried out, focusing on the nature and function of their inputs. Short essays were found to be the standard test input, with most tests employing individual texts as test input and multiple choice as item type, as presented in Table 1 below.

Table 1

Types of test inputs in the existing critical thinking tests

	Name of the Test	Input Types	Item Types	Source
1	California Critical Thinking Skills Test (CCTST)	Short Essays	Multiple Choice	https://www.clemson.edu
2	Cornell Critical Thinking Test (CCTT)	Short Essays	Multiple Choice	www.criticalthinking.com
3	Halpern Critical Thinking Assessment (HCTA)	Short Essays	Multiple Choice	https://marketplace.schuhfried.com
4	Watson–Glaser Critical Thinking Appraisal tool (WGCTA)	Short Essays	Multiple Choice	https://www.assessmentday.co.uk
5	Heighten Critical Thinking Test (HCTT)	Collection of Arguments Short Essays	Multiple Choice	https://www.ets.org

The review of these tests revealed limitations in their utilization of test input. Most of these tests rely heavily on short essays as input and multiple-choice questions as item types. This approach, while convenient, cannot capture the complexity and diversity of complex soft skills such as critical thinking. The exclusive use of essays as test input can be limiting, considering the diverse requirements for assessing the real-world critical thinking abilities of test-takers.

The traditional approach of using short essays as input in testing limits critical thinking assessment to text interpretation and logical reasoning, neglecting dimensions like problem-solving, decision-making, and innovative thinking that are crucial in real-world situations (Staples et al., 2018). Although the HCTT introduced a new input type using a collection of arguments, it still relies on invented literary texts. The need to link varied inputs with their roles in test design has not been adequately addressed in existing tests.

Since critical thinking involves more than text interpretation and logical reasoning, and as critical thinking is a vague concept with varying interpretations (Atkinson, 1997; Ennis, 1992), identifying the most relevant and distinct sub-skills for the proposed MPACT test was a key stage in its design.

2.2.2 Identification of key critical thinking sub-skills

To ensure that the MPACT test effectively measures the most relevant sub-skills of critical thinking, a comprehensive study of the sub-skills of critical thinking was conducted to explore various potential sub-skills. First, an inventory of critical thinking was created. To create the inventory, definitions of critical thinking available in different scholarly works (such as Butler, 2012; Casner-Lotto & Barrington, 2006; Cottrell, 2017; Dewey, 1933; Ennis, 1992; Fisher, 2011; Glaser, 1942; Halpern, 1998; Lai, 2011; Moore, 2013) were examined, and sub-skills inherent in those definitions were listed. In addition, practical sub-skills of critical thinking were also explored based on insights from policy-focused organizations (such as World Economic Forum, UNESCO, OECD), academic institutions (such as University of Tennessee, Rasmussen College), and businesses (such as LinkedIn, Forbes, Indeed).

The initial review of critical thinking sub-skills yielded a lengthy list, many of which were similar in substance but labeled differently. This list was refined to a distinct inventory of 33 sub-skills. To organize these sub-skills, a thematic categorization was conducted. This process analyzed the content and objectives of each sub-skill, identifying four distinct themes as shown in Table 2 below. These themes grouped related sub-skills and addressed various aspects of critical thinking.

Table 2

Thematic categories of critical thinking sub-skills

	Key Themes of Critical Thinking Sub-skills	Number of Sub-skills
1	Numeracy and Likelihood	7
2	Evidence and Analysis	12
3	Reasoning and Argument	6
4	Perspective-taking and Empathy	8

2.2.3 Evaluating criteria of critical thinking sub-skills

The classification of sub-skills into thematic categories streamlined the evaluation process, enabling the identification of key sub-skills within each category. To facilitate the design process of the MPACT test, it was essential to establish criteria for identifying critical sub-skills within those categories. Four main criteria were considered based on the purpose of designing the MPACT test: Innovation, Testability, Clarity, and Real-world Applications. These criteria encompassed the requirements of potential score user organizations for decision-making in areas like hiring, professional development and promotion, while also addressing the limitations of current critical thinking tests. The following section delineates these primary criteria into specific sub-criteria.

2.2.3.1 Innovation

Innovation in the critical thinking sub-skill refers to two main aspects. The first is '*novelty,*' which refers to previously untapped aspects of critical thinking or enhances existing understanding of critical thinking. The second is '*advancement,*' which refers to the potential of the critical thinking sub-skill to bring about meaningful progress or improvements in the field of critical thinking assessment.

2.2.3.2 Testability

Testability in the critical thinking sub-skill encompasses two main aspects. The first is '*measurability,*' which refers to the extent to which a critical thinking sub-skill can be observed, quantified, or assessed. The second is '*objectivity,*' which refers to the degree to which the critical thinking sub-skill can be evaluated using objective test items, such as multiple-choice questions or other closed-ended short-answer questions. These objective test items can be automatically scored to accommodate a large population of test-takers.

2.2.3.3 Clarity

Clarity in the critical thinking sub-skill involves two key elements. The first is '*distinctness,*' which refers to the extent to which the critical thinking sub-skill is clearly defined, easily

understood, and distinguishable from other critical thinking sub-skills. The second element is '*operationalizability,*' which pertains to the degree to which a critical thinking sub-skill can be translated into concrete, observable, and measurable indicators or behaviors that facilitate assessment.

2.2.3.4 Real-world Applications

Real-world applications of the critical thinking sub-skill encompass four key elements. The first is '*contextual relevance,*' which refers to the extent to which the critical thinking sub-skill is applicable to the general needs of the intended target population in which it is intended to be used. The second element is '*real-world problem-solving,*' which pertains to the critical thinking sub-skill's ability to address practical issues and challenges encountered in real-life situations, significantly impacting work and society. The third element is '*transferability,*' which is the ability of the critical thinking sub-skill to be applied across various contexts and situations, demonstrating its versatility and adaptability in addressing diverse real-world problems. The fourth and final element is the critical thinking sub-skills '*alignment with the key principles of industrial-organizational (I/O),*' such as decision-making, leadership, learning and development, and employee well-being in the organizational environment. This alignment ensures that the sub-skill is relevant and contributes to an individual's overall effectiveness and success in the organizational context.

Identifying appropriate main criteria and sub-criteria and their operational definitions was crucial in determining the key sub-skills of critical thinking. This process ultimately resulted in an evaluation framework for identifying the appropriate sub-skills of critical thinking, as presented in Table 3.

Table 3

Evaluation framework for identifying critical thinking sub-skills

	Main Criteria	Sub-Criteria
Critical Thinking Sub-skills	I. Innovation	1. Novelty
		2. Advancement
	II. Testability	1. Measurability
		2. Objectivity
	III. Clarity	1. Distinctness
		2. Operationalizability
	IV. Real-world Applications	1. Contextual Relevance
		2. Real-world Problem-solving
		3. Transferability
		4. I/O Psychology Principles Alignment

Implementing this evaluation framework thoroughly, I examined the sub-skills of critical thinking featured in the inventory. The results of the examination of all sub-skills in relation to each sub-criterion are tabulated in Table 4, represented with the help of a symbol legend. The result symbols and their definitions are as below.

✓ represents the sub-skill meeting the set criterion.
✗ signifies that the sub-skill does NOT meet the set criterion.
? indicates uncertainty regarding whether the sub-skill met the set criterion.

Table 4

Evaluation results of critical thinking sub-skills

Sub-skills of Critical Thinking		Innovation		Testability		Clarity		Real-world Applications			
		1	2	1	2	1	2	1	2	3	4
Theme 1: Numeracy and Likelihood											
1	Be objective in thinking process	✓	✓	✓	✗	?	?	?	?	✓	✗
2	Deconstruct ideas	✓	✗	✓	✓	?	?	✓	?	?	✗
3	Calculate likelihoods (conceptual and numerical both)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
4	Process the information in a creative manner to reach a conclusion	✓	✓	✗	✓	?	?	✓	✓	✓	✓
5	Make a judgement or come to a conclusion based on empathy	✓	✓	✓	✓	?	?	?	?	?	?
6	Make a judgement or come to a conclusion based on culture	✓	✓	✓	✓	?	?	?	?	?	?
7	Reconstruct content or problem	✓	✓	✓	✓	?	?	✓	?	?	✗
Theme 2: Evidence and Analysis											
8	Analyze how parts of a whole interact with each other	✗	✗	✓	✗	✗	✗	?	✓	✓	✓
9	Analyze and evaluate evidence	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
10	Analyze and evaluate claims	✓	✓	✓	✓	?	✓	✓	✓	✓	?
11	Analyze and evaluate beliefs	✓	✓	?	✓	?	?	✓	?	✓	✗
12	Evaluate statistics	✓	✓	✓	✓	?	?	?	✓	✓	✓
13	Evaluate observable phenomenon	✓	✓	✓	✓	?	?	?	?	✓	✗
14	Evaluation of research findings	✓	✓	✓	✓	✓	✓	?	✓	✓	?
15	Interpret information and draw conclusion	✓	✓	✓	✓	?	?	✓	✓	✓	?
16	Generalize the results	✗	✓	✓	✓	✗	✗	?	?	?	✗
17	Formulate inferences	✗	✗	✓	✓	?	?	?	✓	✓	?
18	Identify, analyze and evaluate various opinions	✓	✓	?	✓	?	?	?	✓	✓	✓
19	Find out if something is likely to be true	✗	✓	?	✗	?	?	?	✓	✓	✓
Theme 3: Reasoning and Argument											
20	Use inductive reasoning	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓
21	Use deductive reasoning	✗	✗	✓	✓	✗	✗	✓	✓	✗	?
22	The identification of arguments and non-arguments	✓	✓	✓	✓	?	?	✓	?	?	?
23	Analyze and evaluate arguments	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Sub-skills of Critical Thinking		Innovation		Testability		Clarity		Real-world Applications			
		1	2	1	2	1	2	1	2	3	4
24	Scrutinizing arguments unsupported by logical evidence	✓	✓	✓	?	?	?	✓	?	?	✓
25	Make conclusion that can be defended and justified	✓	✓	✓	✓	?	?	✓	?	?	?
Theme 4: Perspective-taking and Empathy											
26	Identify connections across disciplines	✓	✓	✗	✓	✓	✓	✗	✓	✓	?
27	Challenge the information derived from different sources	✓	✓	?	✗	✓	✓	✓	✗	✓	?
28	Analyze and evaluate major alternative points of view	✓	✓	✓	✓	?	?	✓	✓	✓	?
29	Synthesize and make connections between information and arguments	✓	✓	✓	✓	?	?	✓	?	?	?
30	Analyze one's understanding of the subject from other's perspective	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
31	Justify the ideas and actions	✗	✓	?	✓	?	?	?	✓	✓	?
32	Commit to overcome native egocentrism	✓	✗	?	✓	?	?	?	?	?	?
33	Commit to native sociocentrism	✓	✓	?	✓	?	?	?	?	✗	✗

The validity and reliability of these results were ensured by involving two domain experts who independently reviewed and confirmed the findings. A consensus was reached for disagreements among the test designer and inter-raters.

2.2.4 Selection of the key sub-skills for the MPACT test

The evaluation study revealed that one particular sub-skill emerged as the most suitable within each theme. These sub-skills met all the evaluation criteria and encompassed the essential qualities of the other sub-skills within their respective themes. Consequently, four critical thinking sub-skills were identified as the most suitable, with one originating from each theme. Shorter names were assigned to these key sub-skills to simplify nomenclature and facilitate ease of use. Strong alignment with the evaluation criteria makes them especially suitable for inclusion in the MPACT test, as shown in Table 5 below.

Table 5

Key sub-skills of critical thinking

	Key Sub-skills of Critical Thinking	Innovation	Testability	Clarity	Real World Applications
1	Critical Numeracy (Item 3 in Table 4)	Innovative	Objectively measurable	Unambiguous & operationalizable	Applicable to solving real-world problems
2	Assessing Evidence (Item 9 in Table 4)	Innovative	Objectively measurable	Unambiguous & operationalizable	Applicable to solving real-world problems
3	Argument Evaluation (Item 23 in Table 4)	Innovative	Objectively measurable	Unambiguous & operationalizable	Applicable to solving real-world problems
4	Perspective Taking (Item 30 in Table 4)	Innovative	Objectively measurable	Unambiguous & operationalizable	Applicable to solving real-world problems

2.3 Developing test items

This section explores the intricate process of developing test items for the MPACT test. It chronicles the journey from reviewing existing tests to identifying effective strategies and potential pitfalls. The section further details the process of designing prototype items for the MPACT test, drawing on insights from examining existing critical thinking tests. It also discusses the evaluation of the initial test and the subsequent need for redesigning it to assess critical thinking skills in real-world contexts better.

2.3.1 Reviewing existing tests

Following the identification of the core sub-skills for the MPACT test, a review of existing critical thinking assessments was carried out. Notwithstanding reliability concerns in some tests (Leppa, 1997; Liu et al., 2016; Loo & Thorpe, 1999), the objective was to distill effective strategies from established designs. A recurrent feature was the employment of short passages as test inputs, typically complemented by multiple-choice items. Following is an example from the Heighten Critical Thinking Test (ETS, 2021) which utilizes a short text as test input.

> *"…William Shakespeare of Stratford Could Not Have Written the So-called Shakespearean Plays". We all know that there was a real person named William Shakespeare …"*

Although engaging, this invented abstract of a literary journal as input may not effectively mirror the types of real-world I/O contexts where test-takers typically apply critical thinking skills. This discrepancy in test input could compromise the test's predictive validity, as it fails to assess the skills required in practical, real-life contexts adequately (Staples et al., 2018). The Watson–Glaser Critical Thinking Appraisal (WGCTA) also utilizes short texts as test inputs alongside multiple-choice items (https://www.assessmentday.co.uk/) as shown below.

> *Sarah owns a new company. New companies are more likely to fail than well-established companies. Therefore, Sarah's company will fail.*
> *A. Conclusion Follows*
> *B. Conclusion Does Not Follow*

The multiple-choice item in the WGCTA example, with its oversimplified scenario, not only lacks authenticity in reflecting real-world contexts but also prioritizes formal logic over a nuanced evaluation of critical thinking, underscoring the need for test items that capture the complexities of critical thinking skills within practical scenarios.

2.3.2 Designing prototype items for the MPACT test

Drawing from the insights garnered through the examination of existing critical thinking tests, the groundwork was laid for the design process of the MPACT test items. Given the MPACT test's objective of using automated marking, it was decided to prioritize the multiple-choice format using reading passages.

Short passages (short essays), primarily from the internet, were sought as test inputs for each of the four critical thinking sub-skills. Concurrently, a set of multiple-choice items was generated for each of these sub-skills. Designing tasks and items to assess critical thinking adequately proved challenging. However, after considerable deliberation and several iterations, a collection of prototype items was successfully created for each of the four sub-skills.

2.4 Evaluation of initial test: The failure

Moving forward in the narrative, this section discusses the evaluation phase of the initial MPACT test design. It highlights the feedback received from expert evaluators on the preliminary prototypes of the MPACT test items. This section also presents an example of a prototype item to illustrate these concerns. The narrative then transitions into a discussion on the need for redesigning the test, emphasizing the importance of authenticity and higher-order thinking skills in assessing critical thinking.

2.4.1 Expert evaluation of the MPACT prototypes

The preliminary prototypes of MPACT items across four selected critical thinking sub-skills—Critical Numeracy, Assessing Evidence, Argument Evaluation, and Perspective Taking—were critically evaluated by two experts. They raised two significant concerns: a lack of real-world context diminishing test authenticity and overemphasizing reading comprehension rather than assessing complex critical thinking. These concerns are manifested in the following prototype item designed for the Argument Evaluation sub-skill.

> *…In 2005, Popular Mechanics published a massive investigation of similar claims and responses to them. The reporting team found that the North American Aerospace Defense Command (NORAD) did not have a history of having fighter jets prepped and ready to intercept aircraft that had gone off route...*
>
> *The Popular Mechanics report concludes that…*
> *A. NORAD didn't have a history of having fighter jets.*
> *B. NORAD was not ready to intercept aircraft gone off route.*
> *C. there was no evidence that the government planned the attacks.*
> *D. people blamed the government planned the attacks.*

This item illustrates both concerns: the passage from a Popular Mechanics article may not align with the test-taker's real-world contexts, and the item seems to measure reading comprehension more than Argument Evaluation or application of higher-order thinking skills.

2.4.2 The need for redesigning the test: Areas to explore

As the test designer, I engaged with two experts to address the concerns raised during the evaluation of the MPACT prototype. Our discussions revealed that many of the prototype items predominantly resembled reading comprehension tasks, which raised concerns about their validity in assessing critical thinking skills. After consulting with the experts, it was decided

that these items should go beyond mere comprehension, requiring test-takers to engage in higher-order thinking skills, such as analysis and evaluation, while processing the provided information. By strongly emphasizing authenticity, I decided to embark on a redesign of the MPACT test.

3. Redesigning the MPACT test: Emphasizing authenticity

Recognizing the shortcomings of the initial design led to significant revisions in the test input, task, and item types during the redesign of the MPACT test. Driven by the objective of creating a more authentic measurement of critical thinking and a commitment to improving assessment tools, this redesign aimed to develop a more authentic assessment that reflects real-world critical thinking demands. The journey of the redesign is narrated in seven key steps in the following sections, each discussing a crucial aspect of the redesign and validation process. The strategies for infusing authenticity into every aspect of the test design, from reevaluating critical thinking in the real world to selecting input materials and designing authentic tasks and items.

3.1 Revaluating critical thinking in the real-world

As the MPACT test was reengineered to echo real-world applications of critical thinking, particularly in the workplace, the concept of test authenticity was revisited. Pioneering researchers like Wiggins (1990, 1993) and Maclellan (2004) highlight the importance of test tasks mimicking real-world uses and tasks pertinent to professional contexts. Critical thinking's applicability spans diverse contexts, such as education, work, and life, with specific workplace tasks highlighting professional competency requirements. Thus, the primary objective of redesigning the MPACT test was to align test tasks with real-world challenges that demand critical thinking, enhancing the authenticity of the test. This focus on real-world alignment is essential, as the MPACT test can assist employers in selecting graduates who possess the necessary critical thinking skills for professional success.

3.2 Aligning with current measurement standards

A broader guideline was necessary to enhance the MPACT test and make it a more authentic measure of critical thinking. For this, some emerging practices were reviewed, and I adopted the Standards for Educational and Psychological Testing (APA, AERA, APA & NCME, 2014) to ensure its validity, reliability, and fairness. These standards provide valuable recommendations for test developers, emphasizing the need to explicitly define test constructs and design test content that yields interpretable evidence for intended score interpretations. Moreover, the standards emphasize the importance of carefully integrating test content to ensure its relevance to the measured construct and to mitigate any potential biases while respecting the diversity of test takers.

3.3 Aiming for situational authenticity

The MPACT test redesign specifically emphasized situational authenticity, seeking to emulate real-world scenarios in major test elements such as input, task, item, and response

(Bachman & Palmer, 2022). While psychometric authenticity pertains to score consistency, situational authenticity plays a crucial role in enhancing the predictive validity of workplace performance.

In skill assessments, 'input' refers to the material used as a foundation for tasks and items. 'Tasks' are specific input-linked activities, like distinguishing, categorization, problem-solving, or sequencing. 'Items' are individual questions in various formats, such as multiple-choice, drag-and-drop, or matching. Lastly, 'response' signifies the test taker's answer or solution to an item, revealing their skill level. Authenticity in these elements forms a hierarchy of prerequisites. In other words, if the input is inauthentic, the other elements cannot be authentic. Since it is difficult to achieve situational authenticity in all four elements in a testing context (Watson Todd, 1996), for the MPACT test redesign, I focused primarily on authenticity in the input as a prerequisite but also considered ways to make the other elements more authentic.

The use of authentic input has become common in professional exams such as the AICPA Licensing Examination and Common Core tests. These assessments target specific audiences working in specific contexts and thus it is relatively straightforward to identify relevant authentic input. In contrast, generic tests targeting broad audiences across a range of contexts have very rarely attempted to use authentic input. Recognizing the input's crucial role in representing the construct and content in skill tests, input authenticity was prioritized during the MPACT test's redesign to ensure it aligns with the contexts of test score use.

3.4 Redesigning MPACT with authentic inputs

The focus on input authenticity during the prototype creation for the MPACT redesign process led to selecting specific types of input that favored the test design for various sub-skills of critical thinking. The following sub-sections illustrate one specific type for each sub-skill and their effectiveness in assessing critical thinking.

3.4.1 Tweets as a potential source of test input for perspective taking

The Perspective Taking sub-skill of critical thinking in the MPACT test necessitated innovative input sourcing. Initial attempts with random essays proved insufficient, leading to the exploration of Twitter, a platform rich with diverse viewpoints on varied topics (Keim-Malpass et al., 2017). Tweets, particularly those igniting professional discussions like remote work policies, provided authentic material for assessing Perspective Taking. Below is an example of how tweets were used in the test.

> **Saran:** Been working from home for a year now. Honestly, I'm starting to miss the office vibe. Remote work isn't for everyone!
>
> **Somsri**: I hear you @Saran, but I have a different take. I've found that I'm more productive without the distractions of an office. Remote work can be a real boon for focused tasks.

Kiet: I think it depends on the person and their role. Some jobs just work better in an office, while others are perfect for remote work. Flexibility should be the key.

Siriporn: The challenge is maintaining team coordination when everyone is remote. It's harder to build relationships when you only interact online.

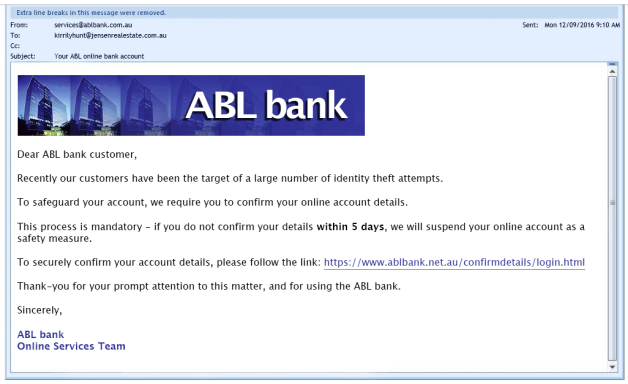
Nattapong: It's all about balance, right? Maybe a mix of remote and in-office work could be the best solution. This hybrid model could bring the best of both worlds.

1. Which individual's tweet implies that a change in management strategies might be needed if a company were to shift completely to remote work?
    A. Saran
    B. Somsri
    C. Kiet
    D. Siriporn
2. Who seems to share a similar view on remote work as Saran, but with a more flexible approach?
    A. Saran
    B. Somsri
    C. Kiet
    D. Nattapong
3. If an organization considers a permanent shift to remote work, which pair of individuals might raise the most concerns based on their perspectives?
    A. Saran and Somsri
    B. Somsri and Kiet
    C. Kiet and Nattapong
    D. Saran and Siriporn

The diversity in tweets facilitated the item-design process for examining others' perspectives (Anderson & Karthwohl, 2001; Dagostino et al., 2014), creating effective higher-order thinking items for Perspective Taking with minor content adjustments.

3.4.2 Online scams as test input for assessing evidence

The MPACT test utilizes scam emails which are a significant global issue (Cross, 2018; PwC, 2020) to assess the sub-skill 'Assessing Evidence' concerning information credibility. Scam and genuine emails were potential input for the task base, requiring test-takers to identify the real and fake emails and provide evidence. Traditional multiple-choice items with one best answer fell short in creating linked items to measure critical thinking. Fourteen pieces of evidence were derived from five emails. A varying number of these pieces of evidence could be used to determine the legitimacy of the respective emails. Thus, a Select-n format (Muckle et al., 2011) was employed, allowing students to decide on the authenticity of the email and select specific pieces of evidence based on their reasoning from multiple correct options as shown in the following example.



(Source: https://www.consumer.vic.gov.au)

1. What do you think about the message? Real or Fake?
    1. Real
    2. Fake

2. While deciding whether this message is real or fake, which of the following pieces of evidence did you use? You can select **one to three** options.

    1. The sender is not available in person.
    2. The sender demands up-front payment.
    3. An online money transfer is required.
    4. There is a demand for urgent action.
    5. Serious consequences are likely if no action is taken.
    6. There is no contact number.
    7. The sender's email ID is linked with the organization's website.
    8. The message contains the telephone number of the customer service center.
    9. Suggested action can be taken either online or in-person.
    10. The sender suggests the steps to get into their company webpage.
    11. The message highlights the long inactivity of the customer.
    12. The sender provides a web link for more detailed info in the message.
    13. The email ID and the domain are different.
    14. There is a demand to access the personal computer.

The inclusion of scams and genuine emails as test input in the MPACT test creates an authentic context for evaluating the information in the workplace based on the evidence. This approach presents practical scenarios that enable test-takers to critically assess information and make evidence-based judgments. Incorporating diverse scam and genuine emails as test input facilitates the design of the tasks mirroring situations commonly encountered in the present technology-integrated professional contexts.

3.4.3 Tables as test input for critical numeracy

In the search for authentic sources of input, tables were explored as a valuable resource to develop test items for the Critical Numeracy sub-skill of critical thinking. Tables are commonly

used across various professions and present numerical data in a realistic format encountered in workplace and organizational contexts. Critical numeracy, a crucial sub-skill of critical thinking, involves interpreting numerical data and applying it to make informed decisions in practical scenarios (Geiger et al., 2015). Utilizing tables as test input provides an authentic evaluation of critical numeracy skills, reflecting the ability to make informed decisions in professional settings. See the example below for the inclusion of table input in the MPACT test.

	Virus disease	Year Identified	Number of Infected Cases	Number of Deaths	Number of Countries with Cases
1	Marberg	1967	1000	800	11
2	Hendra	1994	10	6	1
3	SARS	2002	8,000	720	9

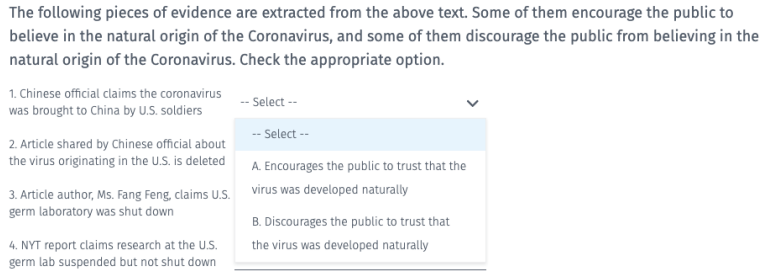
1. Based on the provided table, which virus-infected patient is most likely to survive? Please write only the Virus's name as indicated in the table.

The table input in this closed-ended short-answer question necessitates the interpretation and application of numerical data to make real-world decisions. It evaluates the ability to utilize mathematical knowledge to address practical challenges, such as resource prioritization in organizational settings. The scenario provided highlights the question's practical relevance by requiring interpretation of the table data and informed decision-making. The table serves as a valuable input in crafting the test task, maintaining real-world relevance by calculating the likelihood of future scenarios and demonstrating its applicability across diverse professional and personal contexts.

3.4.4 Conspiracy theories as test input for argument evaluation

The Argument Evaluation sub-skill in the MPACT test uses conspiracy theories as test input, providing an authentic context for assessing critical thinking. This sub-skill includes assessing various argument types, often interlaced with misinformation and fake news, which affect real-world functionality and have economic costs (Cavazos, 2019; Jolley et al., 2020).

A COVID-19 conspiracy theory served as the base for practical critical thinking items, with collected justifications representing diverse theories about the virus's origin. The sample test task is as follows.



This task asked test-takers to weigh the evidence for and against the argument, a skill essential in professional environments where employees often analyze competing proposals and make

evidence-based decisions. Although the test pool included typical workplace examples, COVID-19 conspiracy theories exemplify broader societal relevance.

3.5 The redesign of the full test: A broad table of specification

Identifying suitable types of inputs, tasks, items and responses for each critical thinking sub-skill expedited the construction of the full MPACT test. Real-world input materials like tweets, scam emails, tables, and conspiracy theories were integrated to create authentic tasks that required critical thinking relevant to organizational and industry contexts. Efforts to embed situational authenticity into all test design elements yielded varied outcomes. Near-full authenticity was achieved in the input materials, and partial success was realized with the tasks and items, but limitations were encountered in creating authentic responses due to the objective nature of the MPACT test.

The real-world input integration aligning with each sub-skill of critical thinking was executed to mitigate construct irrelevant variance and potential bias based on the assessment standards for test fairness. A test input was used to create a set of items. Five high-level informants, who were familiar with the contexts and the purpose of the test, reviewed each set of questions to evaluate the relevancy and usability of the input and tasks to the target population, and their feedback on the input content and task authenticity was adjusted. The construct representation process in the MPACT test drew upon Raymond's (2001, 2002) emphasis on capturing essential knowledge, skills, and abilities necessary for a specific domain in future test design. In order to ensure the test focuses on the critical thinking construct rather than language proficiency, a glossary for less common vocabulary was included and identified through corpus analysis of the test input.

The complete test development process showed that the input played a key role in bridging the gap between the critical thinking construct and the MPACT test as a predictor measure. During the development of the complete MPACT test, a basic table of specifications (Table 6) was established.

Table 6

A basic table of specifications for the MPACT test

Critical Thinking Sub-skill	Input Types	Test Tasks	Item Types
Perspective Taking	• Tweets • Collection of short texts from multiple authors • Dialogues	• Identifying perspectives and categorizing supporting and opposing perspectives • Determining the logical order of statements	• Multiple Choice • Sequencing
Assessing Evidence	• Collection of real and spam emails • Collection of statements with facts and opinions	• Distinguishing between reliable and unreliable information • Selecting statements that are supported by evidence	• Real/Fake • Select-n
Critical Numeracy	• Numerical tables • Informative texts with numerical data	• Solving mathematical problems based on given data • Analyzing and interpreting numerical information	• Closed-ended Short Answer

Critical Thinking Sub-skill	Input Types	Test Tasks	Item Types
Argument Evaluation	• Informative texts with multiple arguments	• Identifying the main argument • Categorizing supporting and opposing arguments	• Multiple Choice • Drag & Drop

The redesigned MPACT test underwent review and validation by the same subject-matter experts who evaluated the initial stage prototypes. They incorporated the redesigned test and a scoring scheme that assigned weights ranging from 0 to 3 points to each item, considering the cognitive processing complexity involved. The consensus among the experts is that the redesigned MPACT test aims to serve as a credible tool for assessing individuals' critical thinking proficiency, catering to the needs of employers, educational institutions, and other stakeholders.

3.6 Students' perceptions of the redesigned MPACT test

A perception study on the MPACT test provided valuable insights from the test takers' perspective. To collect data, undergraduate students in English-medium programs at Thai universities were invited to voluntarily take the test online. After completing the test, participants were asked to complete an open-ended questionnaire concerning general perceptions, perceived effectiveness in measuring critical thinking sub-skills, user-friendliness, and encountered problems. The study, which involved 201 undergraduate students, revealed that the MPACT test was useful for evaluating their critical thinking abilities. While some feedback focused on the length and interface of the digital test, most of the comments were positive, highlighting the students' appreciation for the diverse and relevant input materials. The students recognized the value of the test as a learning opportunity and expressed enjoyment in engaging with input such as articles and news. Some of their original comments further illustrate their perspective.

- Great cases were given to the survey takers. I believe these tests can show the level of students' preparedness in facing work environments.

- I love how the texts reflect more of the current society and skill jobs in this period of time by integrating commonly seen Twitter texts, emails, and current issues.

These comments highlight the students' positive perceptions of the MPACT test and their appreciation for the engaging and relevant input materials. The students found value in the test as a means of self-reflection and recognized the real-world applicability of the content presented.

3.7 Psychometric evaluation of the redesigned MPACT test

After examining the test taker’s perceptions of the redesigned MPACT test, it was thoroughly evaluated using Rasch Modeling, affirming its reliability and validity (Dhakal et al., 2023). With an item reliability of 0.96 and an item separation value of 4.84, the test effectively differentiates performance levels. The item fit indices further attest to the test's effectiveness.

The MPACT test's validity, initially supported by Rasch analysis, was further corroborated by a Pearson correlation study. The study revealed moderate correlations among the test's

theoretically identified four sub-skills sections, demonstrating that they should be considered distinct but related aspects of critical thinking, while the strong correlations between each sub-skill and the whole test indicated their collective contribution to the overall construct. These findings, aligned with Taylor's (1990) interpretation criteria, affirmed that the MPACT test effectively measures a unified critical thinking construct, thereby bolstering its validity.

These psychometric findings implied that the integration of diverse inputs, such as tweets, conspiracy theories, scam emails, and tables, contributes to the test's reliability and validity. The comprehensive psychometric study, qualitative scrutiny during the design stages, students' perceptions, and expert evaluation provide various types of information supporting the validity (Brennan, 2006) and reliability of the MPACT test.

4. Conclusion of the narrative: The central role of test input

The MPACT test, exhibiting strong validity and reliability, highlights three key tenets of test design: situational authenticity, skill-specific evaluation, and inclusivity. Enhancing situational authenticity through real-world inputs significantly boosts the test's predictive validity (Embretson & Reise, 2013; Leighton, 2019). Despite this, situational authenticity remains a continuum, as evidenced in the MPACT's varied successes and obstacles across the test's input, task, item, and response elements.

The test aimed for input authenticity using digital content like tweets and scam emails to reflect workplace critical thinking sub-skills. However, maintaining task and item authenticity proved challenging, owing to discrepancies between real-world reactions and test responses and the complex nature of emulating real-world scenarios like scam email detection. Test environment constraints also limited response authenticity. Despite these obstacles, the design process underlined the complexity and importance of situational authenticity, with input authenticity as the central focus, suggesting future research should delve into improving authenticity across all test elements.

The MPACT design also revealed that diverse inputs enhance the evaluation of critical thinking sub-skills and refine item design. Transitioning from traditional short readings to varied, real-world materials also promotes inclusivity and fairness (Solano-Flores, 2023), potentially benefitting minority test-takers (Chong, 2018; Djiwandono, 2006; Noori & Mirhosseini, 2021; Viruru, 2006). However, standardizing the scoring scheme while accommodating diverse inputs posed a challenge. In conclusion, authentic inputs positively influence the test's validity as vital aspects of situational authenticity, necessitating further research to understand their holistic impact on score reliability.

DISCUSSION: TEST INPUT IN SKILL TESTS DESIGN

Test input, as defined at the beginning of this paper, refers to the material or information that test-takers must depend on to answer the questions provided in the test tasks. Particularly in skill tests, which aim to evaluate the ability of test takers to resolve problems using the

provided scenarios, the input is a distinct and critical component. This input can embody diverse forms such as text passages, data sets, audio, visuals, or real-world artifacts, laying the foundation for the test tasks and items aligning with the objectives of the test.

Recognizing test input as an integral part of test design broadens our lens, enabling us to appraise not only the test-taker's responses, but also the quality and relevance of the material they interact with. This comprehensive approach gains increasing importance as the learning sciences pivot from rote learning and recall to skill assessments. While designing the MPACT test, a marked gap in the assessment design literature was realized, highlighting the lack of a comprehensive exploration of the role of test input in skill test design and its influence on test authenticity and effectiveness.

Existing tests often overlook the incorporation of real-world materials despite their importance in enhancing test validity and authenticity. The MPACT test addresses this by using authentic inputs such as diverse tweets, conspiracy theories, scam emails, and quantitative data. These real-world materials in the MPACT test create a more immersive and predictive test context, effectively stimulating higher-order thinking skills compared to conventional short essays. Such inputs enhance situational authenticity and ecological validity by accurately representing scenarios where critical thinking skills are applied (Bachman & Palmer, 1996; Brown et al., 2012; Frey, 2018; Wiggins, 1993; Wu & Stansfield, 2001).

After reviewing input's role in existing tests, the role of input was revisited in key test design textbooks that were initially examined during the early stage of the MPACT test design. As seen in Table 7, the results highlight a prevailing emphasis on item writing with relatively limited attention given to test input based on the space provided to different test design elements. This pattern strengthens the traditional static view of test input in the field, emphasizing the need for a fresh approach to design a new generation of tests.

Table 7

Page allocation for elements in test design textbooks

	Test Design Textbooks	Total number of pages on test development	Number of pages on test writing		Number of pages on other aspects*
			Item writing	Test input	
1	*Measurement and Evaluation in Education and Psychology* by Mehrens and Lehmann (1991)	181	70	0**	111
2	*Language Testing in Practice* by Bachman and Palmer (1996)	168	21	2	145
3	*Designing and Analyzing Language Tests* by Carr (2011)	182	24	1	157
4	*Developing and Validating Test Items* by Haladyna and Rodriguez (2013)	202	43	0**	159
5	Educational Testing and Measurement Kubiszyn and Borich (2013)	106	54	0**	131

*Other aspects: Testing objective, planning, scoring, interpreting scores, etc. components of test development.

**0: Test input was not found anywhere as a topic or subtopic in the given test guides.

Upon reviewing the nature and functions of input in existing critical thinking tests and prominent test design guides, it is apparent that test input is often treated as a fixed element, serving merely as a platform for item design. However, this overlooks test input's significant role, such as enriching situational authenticity of test, facilitating effective skill test design, and catering to diverse test-takers. This highlights the necessity for reimagining the MPACT test approach to test input during test design.

Advancements in educational measurement spotlight the ongoing work in psychometrics and Industrial-Organizational (I-O) psychology to solidify the connection between test content and design (Davidshofer & Murphy, 2005). However, the specific role of test input in skill assessment needs more exploration. Professional tests have started to diversify test stimuli, such as passages, vignettes, and charts (Jang, 2009; Leighton, 2019). For instance, the AICPA Licensing Examination for Accountants uses diverse inputs fitting its clear target audience (Leighton, 2019). These advancements primarily aim to improve content validity through authentic input derived from job or task analysis; however, they do not extend to exploring the role of such input in innovating item design (Dhakal, 2023). Hence, the MPACT test, building on the approach used in such professional tests, demonstrates how diversified inputs can enhance authenticity and validity in designing authentic and diversified item types for general skills.

The input selection criteria of the MPACT test were aligned with real-world tasks corresponding to distinct critical thinking sub-skills to mitigate construct irrelevance and construct underrepresentation concerns (Downing, 2002; Embretson & Reise, 2013). This approach promotes construct representation, elevates authenticity, and ensures cultural responsiveness of assessments. Different input materials favor specific sub-skills of critical thinking: tables foster Critical Numeracy, conspiracy theories highlight Argument Evaluation, scam emails emphasize Assessing Evidence, and tweets promote Perspective Taking. Culturally diverse and globally relevant materials enhance test fairness and minimize cultural bias (Ercikan & Pellegrino, 2017; Sireci & Randall, 2021). However, the complete impact of diversified test inputs, especially in soft skill assessments, remains largely unexplored.

The innovative strength of the MPACT test lies in its dynamic and adaptable approach to test input, using a variety of real-world artifacts to craft complex cognitive items. While authenticity in input was largely achieved, significant challenges emerged in reaching situational authenticity, particularly in test tasks and item types. Moreover, it was particularly difficult to implement authenticity in the response element of test design, especially when constructing an objective test. Other obstacles included sourcing authentic materials, managing input variability, and devising a balanced scoring scheme. These issues were addressed through expert consultations. A further issue is that the input used in MPACT will need to be constantly reviewed and possibly updated to retain authenticity in a changing world where new contexts in which critical thinking is needed may emerge. Future research should delve deeper into these challenges, thus advancing the development of assessments and soft skill testing.

IMPLICATIONS FOR FUTURE TEST DESIGN

Insights gleaned from the reviews of existing critical thinking tests and test design guides, coupled with the first-hand experience during the MPACT test design process, underline the essential role of test input in authentic skill test design. Five key insights hold significant implications for future test design.

1. Prerequisite for designing skill proficiency tests

Test input is usually a prerequisite for creating proficiency skill tests in contrast to some knowledge tests, where all items stand alone. Input creates a context for tasks that mirror organizational contexts and favor certain item types and response elements. Future test designs should, therefore, prioritize the thoughtful selection and integration of test input from the early stages of test development.

2. Improving test authenticity

Authenticity in skill tests can be enhanced by using inputs that simulate real-world scenarios, thereby improving the tests' validity. Input is the cornerstone of situational authenticity, influencing other key test elements like tasks, item types, and responses. The MPACT test, for instance, incorporates real-world materials to create authentic test tasks. Future test designs should consider such inputs to ensure that tasks accurately reflect real-world scenarios and thereby improve the authenticity of skill assessments.

3. Linking criterion construct with predictor measure

The design of the MPACT test highlights the vital role of authentic input in effectively bridging the criterion construct and the predictor measure. In this model, test design elements, including inputs, tasks, items, and responses, collectively form a hierarchical structure aimed at authentically assessing a construct. Notably, while achieving full authenticity with test inputs is relatively straightforward, tasks, items and responses were only partially authentic in the testing context. Thus, input's critical role is underscored, shaping test items that mirror real-world scenarios and demanding the cognitive abilities necessary for critical thinking. As a result, the MPACT test, serving as a predictor measure, is proficient at evaluating the application of these skills in real-world contexts. Importantly, this link between the criterion construct and the predictor measure bolsters the test's predictive validity. Hence, the process reaffirms the importance of incorporating authentic input in creating robust skill assessments, emphasizing the direct applicability of the test results to real-life scenarios.

4. Promoting a pluralistic approach and test fairness

The MPACT test advances pluralism and fairness by incorporating a variety of inputs that are relevant to diverse workplace environments. The integration of such diversified material enhances engagement, mitigates cultural bias, and fosters inclusivity. In this way, the MPACT test serves as an exemplar of how varied and globally pertinent inputs can be utilized in

culturally responsive assessments, particularly when evaluating complex skills such as critical thinking. It underscores the value of representing multiple perspectives in test design and encourages a more equitable approach to assessment.

5. Bridging the assessment-learning gap

The perception study of the MPACT test showed that incorporating diverse, real-world inputs relevant to the workplace and everyday life into assessments enhances test takers' meaningful engagement. It implies that input can contribute to transforming large-scale objective tests into opportunities for learning. Test takers revealed that the input materials made them aware of important issues relevant to their life and future career and made the test interesting and motivating. Students also found unfamiliar input, such as conspiracy theories on the Moon Landing, valuable learning sources. This highlights the importance of future test designs that prioritize integrating authentic, purpose-guided, real-world materials to bridge the gap between assessment and learning.

THE AUTHORS

Khagendra Raj Dhakal is a Ph.D. candidate in applied linguistics at King Mongkut's University of Technology Thonburi, Thailand. He has been teaching applied linguistics and general education courses at King Mongkut's University of Technology North Bangkok for more than a decade. His areas of interest include education policies, assessment development, critical thinking, and global competencies. Dhakal is also a member of the *National Council on Measurement in Education*.
raj.kmutt@gmail.com

Richard Watson Todd, Ph.D., is an associate professor at the School of Liberal Arts, King Mongkut's University of Technology Thonburi, Thailand. He has a Ph.D. from the University of Liverpool, and is the author of numerous articles and several books, including *Discourse Topics* (John Benjamins, 2016). His research interests include text linguistics, corpus linguistics, and educational innovations.
irictodd@kmutt.ac.th

Natjiree Jaturapitakkul, Ph.D., is an assistant professor at the School of Liberal Arts, King Mongkut's University of Technology Thonburi, Thailand, where she teaches English to undergraduate and graduate students in the ELT program. She has published and presented numerous papers on English language teaching and learning, language assessment, test development, and ESP testing.
natjiree.jat@kmutt.ac.th

REFERENCES

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (AERA, APA & NCME). (2014). *Standards for educational and psychological testing*. https://www.testingstandards.net/open-access-files.html

Anderson, L. E., & Karthwohl, D. (Eds.). (2001). *A taxonomy for learning, teaching and assessment*. Longman.

Atkinson, D. (1997). A critical approach to critical thinking in TESOL. *TESOL Quarterly, 31*(1), 71–94. https://doi.org/10.2307/3587975

Bachman, L., & Palmer, A. (2022). *Language assessment in practice: Developing language assessments and justifying their use in the real world.* Oxford University Press.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice designing and developing useful language tests*. Oxford University Press.

Brennan, R. L. (Ed.). (2006). *Educational measurement*. Praeger Publishers.

Brookhart, S. M., & Nitko, A. J. (2014). *Education assessment of students*. Merrill Prentice Hall.

Brown, G. T. L., Harris, L. R., & Harnett, J. (2012). Teacher beliefs about feedback within an assessment for learning environment: Endorsement of improved learning over student well-being. *Teaching and Teacher Education, 28*(7), 968–978. https://doi.org/10.1016/j.tate.2012.05.003

Buranapatana, M. (2006). *Enhancing critical thinking of undergraduate Thai students through dialogic inquiry* [Doctoral dissertation, University of Canberra]. University of Canberra Research Portal. https://doi.org/10.26191/ayj9-rm66

Butler, H. A. (2012). Halpern critical thinking assessment predicts real-world outcomes of critical thinking. *Applied Cognitive Psychology, 25*(5), 721–729. https://doi.org/10.1002/acp.2851

Carr, N. T. (2011). *Designing and analyzing language tests*. Oxford University Press.

Casner-Lotto, J., & Barrington, L. (2006). *Are they really ready to work? Employers' perspectives on the basic knowledge and applied skills of new entrants to the 21st century US workforce*. Partnership for 21st Century Skills. https://files.eric.ed.gov/fulltext/ED519465.pdf

Cavazos, R. (2019). *The economic cost of bad actors on the Internet: Fake news in 2019*. University of Baltimore. https://s3.amazonaws.com/media.mediapost.com/uploads/EconomicCostOfFakeNews.pdf

Chaitrong, W. (2019, October 10). *Lack of critical thinking makes Thailand's competitiveness ranking slip*. The Nation. https://www.nationthailand.com/

Chong, C. S. J. (2018). Battling biases: How can diverse students overcome test bias on the multistate bar examination. *University of Maryland Law Journal of Race, Religion, Gender & Class, 18*(1), 31–97. https://digitalcommons.law.umaryland.edu/rrgc/vol18/iss1/19/

Conley, D. T. (2008). Rethinking college readiness. *New Directions for Higher Education*, (144), 3–13. https://doi.org/10.1002/he.321

Cottrell, S. (2017). *Critical thinking skills: Effective analysis, argument and reflection* (3rd ed.). Bloomsbury Publishing. https://doi.org/10.1057/978-1-137-55052-1

Cross, C. (2018). (Mis)Understanding the impact of online fraud: Implications for victim assistance schemes. *Victims & Offenders, 13*(6), 757-776. https://doi.org/10.1080/15564886.2018.1474154

Cross, C., Richards, K., & Smith, R. (2016). *Improving responses to online fraud victims: An examination of reporting and support*. Criminal Research Grants. https://www.aic.gov.au/sites/default/files/2020-05/29-1314-FinalReport.pdf

Dagostino, L., Carifio, J., Bauer, J. D., Zhao, Q., & Hashim, N. H. (2014). Assessment of a reading comprehension instrument as it relates to cognitive abilities as defined by Bloom's revised taxonomy. *Current Issues in Education, 17*(1), 1-12.

Davidshofer, K. R., & Murphy, C. O. (2005). *Psychological testing: principles and applications* (6th ed.). Pearson.

Dewey, J. (1933). *How we think*. D.C. Heath and Company. https://archive.org/details/dli.ernet.240488

Dhakal, K. R. (2023). *Soft skills for education and work: Developing an innovative test of critical thinking* [Unpublished doctoral thesis]. King Mongkut's University of Technology Thonburi.

Dhakal, K. R., Watson Todd, R., & Jaturapitakkul, N. (2023). Unpacking the nature of critical thinking for educational purposes. *Educational Research and Evaluation, 28*(4–6), 130–151. https://doi.org/10.1080/13803611.2023.2262447

Djiwandono, P. I. (2006). Cultural bias in language testing. *TEFLIN Journal, 17*(1), 81-88. http://dx.doi.org/10.15639/teflinjournal.v17i1/85-93

Downing, S. M. (2002). Threats to the validity of locally developed multiple-choice tests in medical education: Construct-irrelevant variance and construct underrepresentation. *Advances in Health Sciences Education, 7*, 235-241.

Doye, P. (1991). Authenticity in foreign language testing. In S. Anivan (Ed.), *Current developments in language testing* (pp. 103–110). SEAMEO Regional Language Centre. https://files.eric.ed.gov/fulltext/ED350819.pdf

Education Testing Service (ETS). (2021). *Heighten critical thinking sample items*. https://www.ets.org/

Ellerton, P., & Kelly, R. (2022). Creativity and critical thinking. In A. Berry, C. Buntting, D. Corrigan, R. Gunstone & A. Jones (Eds.), *Education in the 21st century: STEM, creativity and critical thinking* (pp. 9–27). Springer International Publishing. https://doi.org/10.1007/978-3-030-85300-6_2

Embretson, S. E., & Reise, S. P. (2013). *Item response theory*. Psychology Press. https://doi.org/10.4324/9781410605269

Ennis, S. (1992). *The generalizability of critical thinking*. Teachers College Press.

Ercikan, K., & Pellegrino, J. W. (2017). Validation of score meaning using examinee response processes for the next generation of assessments. *Validation of score meaning for the next generation of assessments* (pp. 1-8). Routledge. https://doi.org/10.4324/9781315708591-1

Fisher, A. (2011). *Critical thinking: An introduction*. Cambridge University Press.

Frey, B. B. (2018). *Predictive validity, The SAGE encyclopedia of educational research, measurement, and evaluation.* Sage Publications. https://dx.doi.org/10.4135/9781506326139

Galinsky, E. (2010). *Mind in the making: The seven essential life skills every child needs*. Harper Studio.

Geiger, V., Goos, M., & Forgasz, H. (2015). A rich interpretation of numeracy for the 21st century: A survey of the state of the field. *ZDM, 47*(4), 531–548. https://doi.org/10.1007/s11858-015-0708-1

Glaser, E. (1942). An experiment in the development of critical thinking. *Teachers College Record, 43*(5), 409–410.

Gomez, F. (2002). Education as if people matter: A call for critical thinking & humanistic education. *Belizean Studies, 24*(1), 20–37.

Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and validating test items*. Routledge.

Halpern, D. F. (1998). Teaching critical thinking for transfer across domains: Disposition, skills, structure training, and metacognitive monitoring. *American Psychologist, 53*(4), 449–455. https://doi.org/10.1037/0003-066X.53.4.449

Heckman, J. J., & Kautz, T. (2012). Hard evidence on soft skills. *Labour Economics, 19*(4), 451-464.

Hora, M. T. (2019). *Beyond the skills gap: Preparing college students for life and work*. Harvard Education Press. https://doi.org/10.1080/10668926.2018.1488552

Jain, P., & Rogers, M. (2019). Numeracy as critical thinking. *Adults Learning Mathematics, 14*(1), 23-33.

Jang, E. E. (2009). Cognitive diagnostic assessment of L2 reading comprehension ability: Validity arguments for Fusion Model application to LanguEdge assessment. *Language Testing, 26*(1), 31–73.

Jolley, D., Mari, S., & Douglas, K. M. (2020). Consequences of conspiracy theories. In B. Michael & K. Peter (Eds.), *Routledge handbook of conspiracy theories* (pp. 231–241). Routledge. https://doi.org/10.4324/9780429452734-2_7

Keim-Malpass, J., Mitchell, E. M., Sun, E., & Kennedy, C. (2017). Using Twitter to understand public perceptions regarding the# HPV vaccine: Opportunities for public health nurses to engage in social marketing. *Public Health Nursing, 34*(4), 316-323. https://doi.org/10.1111/phn.12318

Kubiszyn, T., & Borich, G. (2013). *Educational testing and measurement* (11th ed.). Wiley Publishing.

Lai, E. R. (2011). Critical thinking: A literature review. *Pearson's Research Reports, 6*(1), 40–41.

Leighton, J. P. (2019). The risk–return trade-off: Performance assessments and cognitive validation of inferences. *British Journal of Educational Psychology, 89*(3), 441–455.

Leppa, C. J. (1997). Standardized measures of critical thinking: Experience with the California Critical Thinking Tests. *Nurse Educator, 22*(5), 29–33.

Lewkowicz, J. A. (2000). Authenticity in language testing: some outstanding questions. *Language Testing, 17*(1), 43–64. https://doi.org/10.1177/026553220001700102

Liu, O. L., Frankel, L., & Roohr, K. C. (2014). Assessing critical thinking in higher education: Current state and directions for next-generation assessment. *ETS Research Report Series, 2014*(1), 1–23. https://doi.org/10.1002/ets2.12009

Liu, O. L., Mao, L., Frankel, L., & Xu, J. (2016). Assessing critical thinking in higher education: the HEIghten approach and preliminary validity evidence. *Assessment & Evaluation in Higher Education, 41*(5), 677–694. https://doi.org/10.1080/02602938.2016.1168358

Loo, R., & Thorpe, K. (1999). A psychometric investigation of scores on the Watson-Glaser critical thinking appraisal new form S. *Educational and Psychological Measurement, 59*(6), 995–1003. https://doi.org/10.1177/00131649921970305

Maclellan, E. (2004). Authenticity in assessment tasks: A heuristic exploration of academics' perceptions. *Higher Education Research & Development, 23*(1), 19–33. https://doi.org/10.1080/0729436032000168478

Majid, S., Liming, Z., Tong, S., & Raihana, S. (2012). Importance of soft skills for education and career success. *International Journal for Cross-Disciplinary Subjects in Education, 2*(2), 1037–1042. https://doi.org/10.20533/IJCDSE.2042.6364.2012.0147

Mehrens, W. A., & Lehmann, I. J. (1991). *Measurement and evaluation in education and psychology* (2nd ed.). Houghton Mifflin Company.

Moore, T. (2013). Critical thinking: Seven definitions in search of a concept. *Studies in Higher Education, 38*(4), 506–522. https://doi.org/10.1080/03075079.2011.586995

Muckle, T. J., Becker, K. A., & Wu, B. (2011, April). *Investigating the multiple answer multiple choice item format* [Paper presentation]. The Annual Meeting of the National Council on Measurement in Education, New Orleans, LA, United States.

Noori, M., & Mirhosseini, S. A. (2021). Testing language, but what?: Examining the carrier content of IELTS preparation materials from a critical perspective. *Language Assessment Quarterly, 18*(4), 382–397. https://doi.org/10.1080/15434303.2021.1883618

Office of the National Education Commission (ONEC). (2003). *National Education Act B.E. 2542 (1999) and amendments (Second National Education Act B.E. 2545 (2002))*. http://www.onesqa.or.th/upload/download/file_697c80087cce7f0f83ce0e2a98205aa3.pdf

Organization for Economic Cooperation and Development (OECD). (2018). *The future of education and skills: Education 2030*. OECD Education Working Papers.

Pathanasethpong, A. (2014, October 6). *Critical thinking? Perish the thought*. Bangkok Post. https://www.bangkokpost.com/opinion/opinion/436130/critical-thinking-perish-the-thought

Pillay, H. (2002). *Teacher development for quality learning: The Thailand education reform project*. Queensland University of Technology.

Ploysangwal, W. (2018). An assessment of critical thinking skills of Thai undergraduate students in private Thai universities in Bangkok through an analytical and critical reading test. *University of the Thai Chamber of Commerce Journal Humanities and Social Sciences, 38*(3), 75–91.

PricewaterhouseCoopers (PwC). (2022). *PwC's global economic crime and fraud survey 2022*. https://www.pwc.com/gx/en/services/forensics/economic-crime-survey.html

Raymond, M. R. (2001). Job analysis and the specification of content for licensure and certification examinations. *Applied Measurement in Education, 14*(4), 369–415.

Raymond, M. R. (2002). A practical guide to practice analysis for credentialing examinations. *Educational Measurement: Issues and Practice, 21*(3), 25–37.

Rios, J. A., Ling, G., Pugh, R., Becker, D., & Bacall, A. (2020). Identifying critical 21st-century skills for workplace success: A content analysis of job advertisements. *Educational Researcher, 49*(2), 80–89. https://doi.org/10.3102/0013189X19890600

Scully, D. (2017). Constructing multiple-choice items to measure higher-order thinking. *Practical Assessment, Research, and Evaluation, 22*(1), Article 4. https://doi.org/10.7275/swgt-rj52

Sireci, S. G., & Randall, J. (2021). Evolving notions of fairness in testing in the United States. In B. E. Clauser & M. B. Bunch (Eds.), *The history of educational measurement: Key advancements in theory, policy, and practice* (pp. 111–135). Routledge. https://doi.org/10.4324/9780367815318-6

Solano-Flores, G. (2023). Response: How serious are we about fairness in testing and how far are we willing to go? *Educational Assessment, 28*(2), 105–117.

Staples, S., Biber, D., & Reppen, R. (2018). Using corpus-based register analysis to explore the authenticity of high-stakes language exams: A register comparison of TOEFL iBT and disciplinary writing tasks. *The Modern Language Journal, 102*(2), 310–332. https://doi.org/10.1111/modl.12465

Suarta, I. M., Suwintana, I. K., Sudhana, I. F. P., & Hariyanti, N. K. D. (2017). Employability skills required by the 21st century workplace: A literature review of labor market demand. *Proceedings of the International Conference on Technology and Vocational Teachers (ICTVT 2017), 102*, 337–342. https://doi.org/10.2991/ictvt-17.2017.58

Taylor, R. (1990). Interpretation of the correlation coefficient: A basic review. *Journal of Diagnostic Medical Sonography, 6*(1), 35–39.

Viruru, R. (2006). Postcolonial technologies of power: Standardized testing and representing diverse young children. *International Journal of Educational Policy, Research, and Practice: Reconceptualizing Childhood Studies, 7*(1), 49–70.

Watson Todd, R. (1996). Can we test listening authentically? *PASAA, 27*, 80–86.

Wiggins, G. (1990). The case for authentic assessment. *Practical Assessment, Research, and Evaluation, 2*(1), 2. https://doi.org/10.7275/ffb1-mm19

Wiggins, G. (1993). Assessment: Authenticity, context, and validity. *Phi Delta Kappan, 75*(3), 200–213.

World Economic Forum (WEF) (2013). Digital wildfires in a hyperconnected world. *Global Risks 2013* (pp. 23–27). https://www3.weforum.org/docs/WEF_GlobalRisks_Report_2013.pdf

Wu, W. M., & Stansfield, C. W. (2001). Towards authenticity of task in test development. *Language Testing, 18*(2), 187–206. https://doi.org/10.1177/026553220101800205