

Designing a contextualized PIAAC assessment environment for Indonesian adults: An educational design research study

Zulkardi^{1*}, Ratu Ilma Indra Putri¹, Ahmad Fauzan², Laela Sagita³, Rahma Siska Utari¹, Viona Adelia⁴, Yovika Sukma¹, Dea Alvionita Azka¹, Dhea Anisah Putri¹, Nadya Fitriani¹, Novita Sari¹, Rizha Alfajri⁴

¹Department of Mathematics Education, Universitas Sriwijaya, South Sumatra, Indonesia

²Department of Mathematics, Universitas Negeri Padang, West Sumatra, Indonesia

³Department of Mathematics Education, Universitas PGRI Yogyakarta, Yogyakarta, Indonesia

⁴Center of Excellence of PMRI and PISA, Universitas Sriwijaya, South Sumatra, Indonesia

*Correspondence: zulkardi@unsri.ac.id

Received: Jul 12, 2026 | Revised: Sep 12, 2026 | Accepted: Sep 26, 2026 | Published Online: Oct 8, 2026

Abstract

The Programme for the International Assessment of Adult Competencies (PIAAC) assesses adults' literacy, numeracy, and problem-solving competencies in functional situations. However, PIAAC-oriented assessment resources that reflect Indonesian adult contexts remain limited. This study aimed to design and evaluate a contextualized PIAAC assessment environment for Indonesian adults. Educational Design Research was employed using Plomp's preliminary research, prototyping, and assessment phases, integrated with formative evaluation through focus group discussions, one-to-one evaluations, small-group evaluations, and field testing. The study produced 93 contextualized items comprising 32 literacy, 31 numeracy, and 30 problem-solving items, which were incorporated into a computer-based testing system and disseminated through an item book and social media content. Iterative evaluation identified and addressed issues related to contextual relevance, competency alignment, instructional clarity, response-format flexibility, and technical functionality. A field test involving 30 teachers showed mean achievement percentages of 84.33% in literacy, 80.33% in problem solving, and 74.33% in numeracy. The findings indicate that international adult competency constructs can be operationalized within authentic Indonesian contexts and iteratively transformed into a practical, usable digital assessment environment. The study offers design-based guidance for developing culturally relevant adult competency assessments while maintaining alignment with the PIAAC framework.

Keywords:

Collaboration, Design research, Development research, Mathematical ability, Quality education

How to Cite:

Zulkardi, Z., Putri, R. I. I., Fauzan, A., Sagita, L., Utari, R. S., Adelia, V., Sukma, Y., Azka, D. A., Putri, D. A., Fitriani, N., Sari, N., & Alfajri, R. (2026). Designing a contextualized PIAAC assessment environment for Indonesian adults: An educational design research study. *Infinity Journal*, 15(4), 1105-1134. <https://doi.org/10.22460/infinity.v15i4.p1105-1134>

This is an open access article under the [CC BY-SA](#) license.



1. INTRODUCTION

Rapid technological, economic, and social transformations have fundamentally changed the competencies adults need to participate effectively in contemporary life (Díaz-García et al., 2022). The expansion of digital technologies has increased the volume, speed, and diversity of information that adults encounter, while digitalization has changed how information is accessed, communicated, and used in both work and everyday contexts. Adults are therefore increasingly expected to interpret multimodal information, evaluate quantitative evidence, navigate digital systems, and make decisions in situations characterized by changing information and multiple constraints (Yamashita et al., 2019). These changes are also reshaping the workplace, where many tasks increasingly require workers to process information, communicate effectively, and adapt changing conditions, and solve non-routine problems alongside occupation-specific knowledge (Carnevale & Smith, 2013). Consequently, literacy, numeracy, and problem solving are becoming increasingly important not only for employment but also for adults' capacity to make informed decisions in everyday context, including accessing and using information related to health and other aspects of daily life (Yamashita et al., 2019). Moreover, persistent low levels of these foundational competencies may contribute to cumulative social and employment disadvantages, particularly among young adults who experience prolonged disconnection from education and employment (van Vugt et al., 2024). These developments underscore the need for assessment approaches that capture how adults actually mobilize competencies in functional situations rather than merely measuring the retention of school-based knowledge.

The Programme for the International Assessment of Adult Competencies (PIAAC) was developed in response to this changing conception of adult competence and the growing need to assess how adults apply skills in real-life situations. PIAAC represents a large-scale international effort to examine how adults use information-processing skills across personal, social, and work-related situations (OECD, 2013; Schleicher, 2008). Its assessment framework focuses on literacy, numeracy, and problem solving in technology-rich environments, with subsequent developments increasingly emphasizing adaptive problem solving in dynamic and changing contexts (Kirsch & Lennon, 2017; OECD, 2024). This orientation is particularly relevant to contemporary adult competence because PIAAC does not treat these skills simply as bodies of knowledge to be recalled. The design of PIAAC is distinctive because assessment tasks are constructed around the interaction between cognitive processes, task characteristics, contexts, and information representations rather than around conventional disciplinary content alone (Kirsch et al., 2020). Research examining PIAAC through cognitive taxonomies has similarly demonstrated that its tasks encompass different levels and forms of cognitive engagement, highlighting the importance of understanding what respondents are required to do with information rather than simply whether an answer is correct (Boeren & Íñiguez-Berrozpe, 2022). Accordingly, PIAAC provides an important conceptual and methodological basis for assessing adult competencies as functional capacities that are mobilized in response to contextual demands, rather than as the mere retention of school-based knowledge.

A central implication of this approach is the importance of authenticity and context in assessment. Authentic assessment requires meaningful correspondence between assessment activities and the ways knowledge and skills are applied beyond formal testing situations.

Critical elements include task realism, cognitive challenge, the use of knowledge in context, and opportunities to demonstrate performance through meaningful actions (Ashford-Rowe et al., 2014). This orientation is highly compatible with adult competency assessment because adults encounter literacy, numeracy, and problem-solving demands through activities such as interpreting public information, comparing prices, managing schedules, evaluating financial options, reading workplace documents, and navigating digital services. Studies using PIAAC data have shown that adult competencies are closely intertwined with educational, occupational, and social experiences (Kim, 2018; Lopes et al., 2020). From this perspective, the authenticity of an assessment context is not merely a decorative feature of an item; rather, it shapes the nature and complexity of the cognitive processes through which competence is demonstrated (Abbasi, 2025; Makeham & Lee, 2012; Villarroel et al., 2018). In authentic tasks, respondents must determine which information is relevant, select an appropriate strategy, and apply competencies within the constraints of a particular situation (Makeham & Lee, 2012; Oroh et al., 2025; Villarroel et al., 2018).

However, the use of international adult competency frameworks raises an important contextual challenge. Literacy practices are socially and culturally situated, and the meaning of competent performance cannot be fully separated from the contexts in which adults encounter and use information. Perry et al. (2020) have critically argued that large-scale literacy assessments may insufficiently represent the diversity and situated nature of everyday literacy practices. Differences in educational pathways and opportunities can also produce substantially different experiences of skill development among adult populations (Cosser, 2018), while cross-national and intergenerational variations demonstrate that literacy gaps are embedded within broader social structures (Kim, 2018). These concerns are particularly relevant to Indonesia, where improving adult competencies and aligning skills with changing labour-market demands remain significant policy priorities (OECD, 2021). Although PIAAC offers a robust international framework, assessment tasks developed for international comparison cannot simply be treated as ready-made learning or assessment resources for Indonesian adults. There is a need to translate the underlying competency constructs into authentic tasks that preserve the cognitive characteristics of PIAAC while drawing on information, services, occupations, and situations recognizable within Indonesian adult life.

The challenge becomes more complex when problem solving is situated in technology-rich environments. Digital problem solving involves more than technical proficiency; performance depends on how individuals search for information, sequence actions, evaluate alternatives, and modify strategies as a task unfolds. Vörös and Rouet (2016) demonstrated that strategy use is meaningfully related to performance in digital problem-solving tasks, while process-data research has revealed that sequences of actions can provide important evidence about how respondents approach PIAAC problems (Goldhammer et al., 2020; He & von Davier, 2015). Sequential patterns in problem-solving environments have also been associated with adults' background and employment characteristics (Liao et al., 2019). Similar concerns have emerged in professional contexts, where adults, including teachers, require opportunities to develop problem-solving competencies in technology-rich environments through meaningful professional and workplace learning experiences (De Wever et al., 2023). Consequently, a PIAAC-oriented assessment environment should not merely transfer paper-

based questions to a computer screen. Rather, it should integrate authentic stimuli, multiple representations, flexible response formats, and digitally mediated situations that require users to access, interpret, apply, evaluate, and adapt information. In this study, a contextualized PIAAC assessment environment refers to an integrated system comprising contextualized PIAAC-oriented assessment items, a computer-based testing (CBT) system, an item book, and complementary dissemination media. The assessment items constitute the core assessment content; the CBT system provides the digital interface for presenting stimuli, recording responses, and delivering the assessment; the item book provides a supporting non-digital representation of the contextualized tasks; and dissemination media extend access to information about the assessment and adult competencies beyond the formal testing setting. Together, these components constitute the assessment environment developed and evaluated through the EDR process.

Despite extensive international research using PIAAC, much of the literature has concentrated on secondary analysis of large-scale datasets. Researchers have examined international competency patterns, sociodemographic differences, motivation-to-learn measurement, employment characteristics, and methodological procedures for analysing PIAAC data (Gorges et al., 2017; Jakubowski & Pokropek, 2019; Rammstedt et al., 2015). Other studies have advanced process-data analysis by investigating log files and action sequences generated during computer-based problem solving (Goldhammer et al., 2020; He & von Davier, 2015). These studies have substantially strengthened understanding of adult competencies and assessment behaviour. Nevertheless, comparatively limited attention has been given to the design process through which the PIAAC framework can be transformed into locally contextualized assessment items and an accessible digital assessment environment for adult learners, particularly in Indonesia. This constitutes a design-oriented research gap: existing evidence explains what PIAAC measures and how its data can be analysed, but provides less empirical guidance concerning how contextualized PIAAC-oriented assessment resources can be systematically designed, iteratively evaluated, and implemented in a specific national and cultural setting. Beyond the absence of contextualized assessment resources, there is also limited knowledge regarding the design principles required to translate an international competency framework into a coherent digital assessment environment for adult learners. Developing such an environment involves more than adapting item content; it requires aligning competency constructs, authentic contexts, cognitive processes, digital interfaces, response formats, and performance evidence into a coherent assessment system. Consequently, the challenge is not only whether contextualized PIAAC tasks can be produced, but also how they should be designed so that the intended competencies remain theoretically valid while becoming meaningful and accessible within local cultural settings.

Addressing this gap requires systematic development rather than direct translation or superficial modification of existing tasks. Robust literature synthesis is necessary to identify the theoretical and empirical principles underlying assessment design (García-Peñalvo, 2022; Turk, 2021), but theoretical alignment alone cannot ensure that an item is understandable, contextually meaningful, or technically functional for its intended users. The difficulty of an assessment task is also multidimensional and may be influenced by cognitive operations, the complexity of information, and the relationships respondents must construct when generating

a response (Chen & Nunes, 2025). Therefore, locally contextualized PIAAC items need to be examined in terms of competency domain, cognitive process, context, stimulus, item format, expected response, and estimated difficulty. They also require iterative evaluation involving experts and prospective users. Such a process is particularly important for computer-based assessment because apparently minor issues—including ambiguous instructions, rigid answer validation, obscured visual information, or inconsistent response formats—may interfere with respondents' performance and confound the competencies intended to be measured. Educational Design Research (EDR) offers an appropriate methodological approach to address this challenge because it combines systematic product development with theory generation through iterative design, evaluation, and refinement (McKenney & Reeves, 2025; Plomp, 2013). Rather than treating assessment development as a linear production process, EDR enables researchers to investigate how theoretical principles are operationalized into practical design decisions while continuously incorporating evidence from experts and prospective users. Through successive cycles of formative evaluation, the resulting assessment environment is expected not only to satisfy criteria of validity and usability but also to generate empirically grounded and potentially transferable design principles for contextualizing international competency assessments.

Furthermore, adult competency assessment has potential educational functions beyond producing test scores. PIAAC research has generated important evidence concerning the relationship between adult skills, participation, employment, and sustainable development (OECD, 2016). Contextualized assessment tasks can make literacy, numeracy, and adaptive problem-solving demands more visible and recognizable to adults, educators, and other stakeholders. In this study, the contextualized tasks were also organized into an item book and complementary social media content to extend access to competency-related information beyond the formal assessment setting.

Accordingly, this study extends existing PIAAC research in three important ways. First, it shifts the focus from analysing assessment outcomes toward designing assessment environments grounded in authentic Indonesian contexts. Second, it demonstrates how literacy, numeracy, and adaptive problem-solving constructs can be systematically operationalized into contextualized digital assessment tasks through an iterative Educational Design Research process. Third, beyond producing a functional assessment platform, the study generates empirically informed design principles for contextualizing an international adult competency framework while preserving its underlying competency constructs. The contextualized items, computer-based testing system, item book, and dissemination materials therefore represent practical manifestations of the resulting design knowledge rather than the primary contribution of the study. These principles may inform future efforts to contextualize international competency frameworks in other adult education settings.

Against this background, the present study aimed to develop and evaluate a contextualized PIAAC assessment environment for adults in Indonesia and, through this iterative Educational Design Research process, to generate empirically informed design principles for contextualizing an international adult competency assessment framework. Using Educational Design Research through the Plomp development phases and iterative formative evaluation, the study developed PIAAC-oriented items across literacy, numeracy, and

problem-solving domains; integrated the items into a computer-based assessment environment; and refined the products through collaborative review, one-to-one evaluation, small-group evaluation, and field testing. The study also examined the dissemination of contextualized PIAAC tasks through a structured item book and social media content. Accordingly, this study addresses the following research question: How can a contextualized PIAAC assessment environment be systematically designed and iteratively evaluated to produce valid, practical, and usability literacy, numeracy, and problem-solving tasks for adults in Indonesia?

2. METHOD

This study employed Educational Design Research (EDR) to develop a web-based learning environment that supports the development of adults' competencies assessed in the Programme for the International Assessment of Adult Competencies (PIAAC). Educational Design Research is an intervention-oriented research approach that aims to simultaneously address authentic educational problems while generating practical solutions and transferable design knowledge through iterative cycles of analysis, design, evaluation, and refinement (McKenney & Reeves, 2025). Unlike conventional product development studies, EDR emphasizes continuous interaction between theoretical understanding and practical implementation, allowing the developed intervention to evolve based on empirical evidence collected throughout the research process (McKenney & Reeves, 2025).

The development process followed the Plomp development model, which consists of three interconnected phases: Preliminary Research, Prototyping Phase, and Assessment Phase. The preliminary research phase focused on identifying educational needs, analysing the PIAAC framework, and establishing design principles for the learning environment. The prototyping phase involved iterative development and revision of the prototype through Tessmer's formative evaluation, including self-evaluation, expert review, one-to-one evaluation, and small-group evaluation. Finally, the assessment phase evaluated the quality of the final product in terms of validity, practicality, and initial feasibility through field implementation. The integration of Plomp and Tessmer enables systematic product refinement while ensuring that the resulting learning environment is theoretically grounded and empirically validated.

The primary product developed in this study was a web-based PIAAC learning environment that integrates locally contextualized PIAAC assessment items, a digital learning platform, printed learning resources, and social media learning content to facilitate adult learning.

2.1. Participants

Participants were selected purposively according to the objectives of each development stage. During the preliminary research phase, several lecturers, mathematics educators, and educational practitioners participated in interviews and discussions to identify the characteristics of adult learning, existing assessment practices, and the need for contextualized PIAAC assessment resources. The formative evaluation involved different participant groups.

Expert review was conducted by specialists in mathematics education, educational assessment, educational technology, and language to examine the content validity, construct validity, usability, and linguistic appropriateness of the developed assessment resources. One-to-one evaluation involved six teachers to identify potential problems in item comprehension, stimulus clarity, response processes, and interface usability. Small-group evaluation involved six prospective teachers to examine the practicality, clarity, and usability of the contextualized assessment tasks and digital assessment environment. Finally, the field test involved 30 teachers to obtain initial evidence regarding the feasibility and usability of the developed assessment environment in a broader assessment setting. The use of teachers and prospective teachers at these stages was based on practical considerations and the assumption that teachers play an important role in supporting adult education and competency development. These participants served as development and evaluation participants rather than as representatives of the broader target population of Indonesian adults, who are the intended end users of the contextualized PIAAC assessment environment.

2.2. Research Procedure

The research procedure followed the three phases of the Plomp development model integrated with Tessmer's formative evaluation, as illustrated in Figure 1.

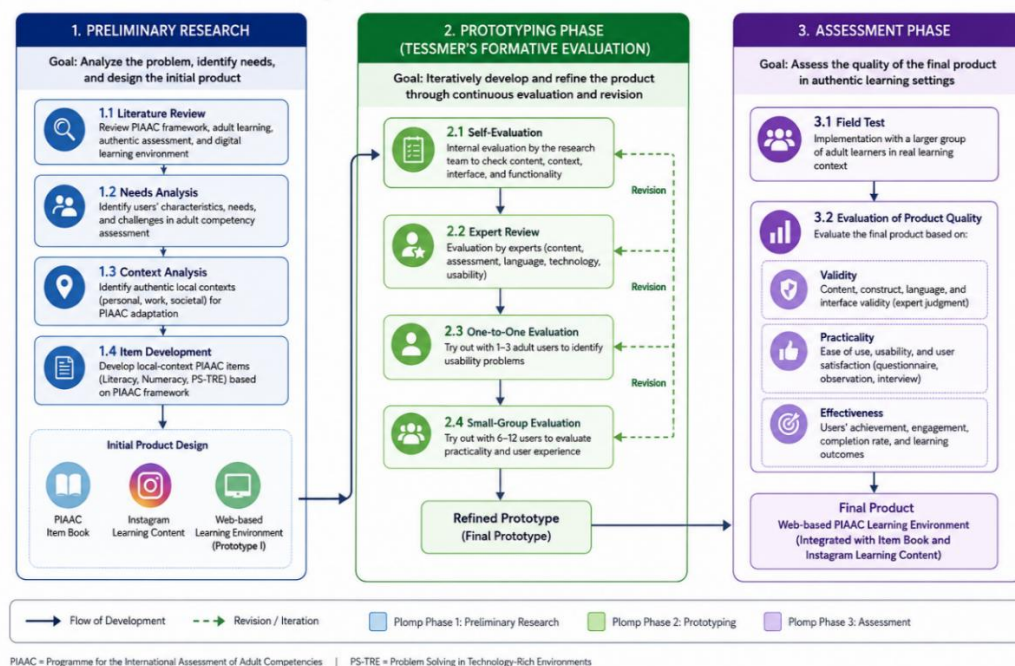


Figure 1. Research procedure based on Plomp development model integrated with Tessmer's formative evaluation

2.2.1. Preliminary research

The preliminary research phase aimed to establish the theoretical and contextual foundation for developing the web-based PIAAC learning environment. Four complementary activities were conducted during this phase. First, a literature review was conducted to examine the OECD PIAAC framework, adult learning theory, authentic assessment, technology-rich

learning environments, and prior studies on PIAAC item development. The findings were synthesized into a set of preliminary design principles guiding the development of contextual assessment items and digital learning features.

Second, a needs analysis was conducted to identify the requirements of prospective users regarding adult competency assessment, digital learning environments, and contextual learning materials. Information gathered from lecturers, teachers, and educational practitioners was used to determine the essential characteristics and functionalities of the proposed learning environment. Third, a contextual analysis was conducted to identify authentic situations that represent Indonesian adult life. The analysis considered personal, workplace, and societal contexts commonly encountered by adults to ensure that the developed items reflected meaningful real-life situations while maintaining alignment with the PIAAC assessment framework.

Finally, based on these analyses, locally contextualized PIAAC items were developed covering the three competency domains assessed by PIAAC: Literacy, Numeracy, and Problem Solving in Technology-Rich Environments (PS-TRE). Each item was documented using a standardized structure consisting of competency domain, cognitive process, context, stimulus, item format, answer key, and estimated difficulty level. The resulting items were subsequently transformed into three integrated learning products: (1) a printed PIAAC Item Book, (2) Instagram-based learning content for public dissemination and informal learning, and (3) Prototype I of the web-based PIAAC learning environment.

2.2.2. Prototyping Phase

The prototyping phase adopted Tessmer's formative evaluation to facilitate iterative refinement of the developed learning environment. Formative evaluation is intended to identify potential weaknesses during development so that revisions can be made continuously before broader implementation (Elwy et al., 2020).

The first activity was self-evaluation, in which the research team examined the prototype internally. The review focused on consistency with the PIAAC framework, contextual relevance, interface design, navigation, item accuracy, scoring procedures, and system functionality. Necessary revisions produced Prototype II.

The second activity was expert review. Experts evaluated the content validity, construct validity, contextual appropriateness, interface usability, language quality, and technical functionality of the prototype using structured validation instruments. Suggestions and recommendations from experts were incorporated into the subsequent prototype revision.

The third activity was one-to-one evaluation, during which three representative adult users individually interacted with the website while being observed by the researchers. Participants were encouraged to express difficulties encountered during navigation, interpretation of instructions, completion of assessment tasks, and understanding of feedback. The observations and interviews provided qualitative evidence for improving system usability.

The fourth activity was small-group evaluation, involving a group of adult participants who completed learning activities using the developed platform. This stage examined the practicality of the learning environment, including ease of use, clarity of learning activities,

interface consistency, navigation, system responsiveness, and overall learning experience. Feedback obtained during this stage informed the final revision before field implementation.

2.2.3. Assessment Phase

The assessment phase aimed to evaluate the overall quality of the final web-based PIAAC learning environment under authentic learning conditions. The final prototype was implemented with a larger group of adult learners representing the intended users of the system. The evaluation focused on three quality criteria commonly adopted in educational design research. Validity referred to the appropriateness of the content, construct, language, and interface of the developed learning environment based on expert judgment. Practicality referred to the degree to which participants could easily use the learning environment during learning activities. Practicality was examined through user questionnaires, observations, interviews, and system usage records. Initial feasibility referred to the ability of the developed learning environment to facilitate participants' engagement and competency development. Initial feasibility was examined by analysing participants' performance on PIAAC tasks, completion rates, learning engagement, and users' overall responses after using the platform.

2.3. Data Collection Instruments

Multiple instruments were employed throughout the development process to collect both qualitative and quantitative data. [Table 1](#) summarizes the instruments used in each research phase.

Table 1. Data collection instruments

Phase	Data Source	Instrument	Purpose
Preliminary Research	Lecturers, teachers, practitioners	Semi-structured interview	Needs analysis
Preliminary Research	Literature	Document analysis	Derive design principles
Expert Review	Experts	Validation sheet	Assess content and construct validity
One-to-One	Adult participants	Observation guide and interview	Identify usability problems
Small Group	Adult participants	Practicality questionnaire	Evaluate usability and practicality
Field Test	Adult learners	Achievement test, questionnaire, observation	Evaluate effectiveness

2.4. Data Analysis

Qualitative data obtained from interviews, observations, expert comments, and user feedback were analysed using thematic analysis to identify recurring issues and formulate revisions for subsequent prototypes. The analysis emphasized identifying design improvements related to contextual authenticity, usability, learning activities, and system functionality.

Quantitative data were analysed descriptively. Content validity was determined based on expert judgments using Aiken's V coefficient. Practicality was calculated as the percentage of positive responses obtained from users during one-to-one, small-group, and field evaluations. Effectiveness was analysed descriptively based on participants' achievement, completion rates, and learning engagement throughout the implementation process.

3. RESULTS AND DISCUSSION

3.1. Results

3.1.1. Preliminary Phase

PIAAC Item Development

Drawing on the literature review conducted prior to this study, we formulated several implications for designing and adapting PIAAC items. Table 2 presents the implications.

Table 2. Implications for PIAAC item development

Aspect of the PIAAC Framework	Synthesis of Literature Findings	Implications for Item Adaptation/ Design
Literacy	The ability to access, understand, evaluate, and reflect on written texts.	Items should include authentic texts, such as announcements, instructions, forms, short articles, or digital information.
Numeracy	The ability to use and reason with mathematical information in various representations.	Items should involve data, tables, graphs, measurement, estimation, percentages, probability, or quantitative relationships in real-life contexts.
Adaptive Problem Solving/ PS-TRE	The ability to solve dynamic problems, particularly in technology-rich environments.	Items should require information searching, strategy selection, evaluation of alternatives, and adjustment of solutions.
Context	PIAAC uses personal, work-related, social, and community contexts.	Item adaptation should be closely related to the lives of lecturers, teachers, education personnel, or the targeted adult participants.
Cognitive Process	Includes information processing, decision-making, and strategy use.	Item design should assess thinking processes, not merely final answers.
Task Characteristics	Uses authentic, information-based tasks that may involve technology.	Item stimuli may include texts, tables, images, application displays, forms, schedules, or digital situations.

The review showed that PIAAC does not merely assess isolated knowledge or school-based skills but measures adults' ability to apply literacy, numeracy, and problem-solving competencies in functional, context-dependent situations (Zulkardi et al., 2026). Therefore, the development of PIAAC items in this study was positioned as an essential component of the assessment environment created to support adults' competencies needed in daily life.

In this environment, items were not treated as conventional test questions, but as contextual tasks that invite adults to interpret information, make decisions, apply mathematical reasoning, evaluate alternatives, and respond adaptively to real-life demands. Therefore, PIAAC items need to reflect real-world situations and require individuals to apply problem-solving skills, such as organizing schedules, planning activities, or resolving technical problems (Zulkardi et al., 2026). Sagita et al. (2026) further elaborated the numeracy domain into specific situational applications that reflect everyday numeracy practices, including daily transactions involving prices, budgeting, money handling, price comparison, discounts and promotions, savings and accumulation; time and scheduling; utility usage; stock and inventory models; reading tables and interpreting graphs. The items were drafted using the adaptation and redesign structure presented in Table 3. Subsequently, Table 4 presents the distribution of the developed items.

Table 3. Structure for item adaptation and redesign

Unit Name – Item #	Item code or identity, for example: LIT01, NUM02, PSTRE03
Process	Type of cognitive process being measured
Source	Type of stimulus used
Context	Situation used in the item
Item Format	Item type, such as multiple choice, short answer, or drag and drop (ICT-based)
Answer	Answer key or expected response
Estimated Difficulty	Level of difficulty

Table 4. Summary of item structures

Domain/ Unit	Item Code	Number of Items	Process	Stimulus	Context	Item Format	Estimated Difficulty
Numeracy	NUM01 – NUM31	31	Access and assess situations mathematically; act on/use mathematics; evaluate and make judgments	Floor plans, fare tables, graphs, infographics, financial data, menus, product images, production tables, transportation data	Work, Personal, Societal/Community	Number entry, single-selection multiple choice, complex multiple choice, short answer	Levels 2–5; predominantly Levels 3–4
Literacy	LIT01– LIT32	32	Accessing text, understanding, evaluating, reflecting	Informational texts, digital services, announcements, procedures, short articles, health/social information	Personal, Societal/Community, Work	Multiple choice, short answer, information identification	Level 1– 5
PS-TRE	PSTRE 01– PSTRE 30	30	Defining, searching, application, evaluating operators, strategy use, adaptive decision-making	Schedules, capacity tables, production data, changing situations, limited-resource scenarios, menus, budgets, public services	Work, Personal, Societal/Community	Multiple choice, multiple selection, number entry, short answer, strategic response	Low– Difficult; predominantly medium

The development results show that the designed items covered personal, work, and societal/community contexts and included PIAAC attributes such as cognitive processes, content areas, contexts, item formats, and estimated difficulty levels. In the numeracy domain, for example, items were developed on topics such as household tile requirements, laundry business layout planning, planning a teacher's side business, training costs, and data interpretation. In the problem-solving domain, items were also developed to require strategy adjustments in dynamic situations, such as scheduling community events, managing resources, and making decisions under constraints.

3.1.2. Prototyping Phase

Focus Group Discussion (FGD)

After the initial PIAAC items were drafted, the first stage of the prototyping process was conducted via FGD. This process involved a review by the research team, during which the developed items were collaboratively checked and cross-reviewed among team members. The purpose of this stage was to examine whether the items were conceptually aligned with the PIAAC framework and sufficiently clear, contextual, and appropriate for adult participants.

The FGD focused on three main aspects: content accuracy, clarity of the stimulus, and relevance of the context to adults' daily lives. Through this process, the team identified several aspects that needed improvement. These included inaccurate or overly general cognitive-process labels, and mismatches between the item's context and the competency domain being assessed. Some items also required clearer answer keys, especially those involving open responses or adaptive decision-making.

a. Content Accuracy

A well-designed survey is essential for achieving accurate and comparable test scores across countries and languages. This implies that content accuracy is closely tied to the quality of item development and the overall survey framework (Kirsch et al., 2020). The indicators of the content accuracy aspect include content relevance, content comprehensiveness, and competency alignment (Mokkink et al., 2025). The experts suggested replacing an unfamiliar foreign transportation context with a locally familiar transportation setting. However, they emphasized that the item should retain the requirement to interpret tabular information, compare alternatives, and make a quantitative decision.

b. Stimulus Clarity

Stimulus clarity concerns whether the information presented in an item can be understood by respondents as intended. Instructions and item objectives should therefore be explicit and sufficiently detailed so that respondents understand what is required and can provide the intended response (Leighton et al., 2011). In a contextualized PIAAC assessment, clarity also requires that the language, visual information, prompts, and contextual cues are familiar and unambiguous for Indonesian adults without providing unintended clues or altering the cognitive demand of the item. Therefore, the items should undergo expert review and content validation to identify potentially ambiguous wording, unfamiliar expressions, misleading visual cues, or unclear instructions. In this study, these issues were examined

during the FGD, in which experts evaluated whether the stimulus and instructions could be interpreted consistently by the target population (Venancio et al., 2020).

The revisions were intended to improve stimulus clarity rather than simplify the assessment. In particular, ambiguous wording, unnecessary information, and unclear visual or contextual cues were revised because these features could introduce construct-irrelevant difficulty. The intended cognitive demand of each PIAAC domain was retained.

c. Context Relevance

Context relevance refers to the extent to which the situations represented in the assessment reflect meaningful and authentic real-world experiences of the target population. This includes examining the real-world relevance, authenticity, and transferability of knowledge embedded in the test items (Ruiz-Primo & Li, 2016). The contextual features selected for each item should accurately represent situations that Indonesian adults may encounter in their personal, occupational, and social lives. The assessment should also include a range of contextual demands, from relatively familiar and straightforward situations to more complex situations requiring respondents to interpret information, make decisions, and apply their knowledge in realistic contexts (González-Such et al., 2016).

The context relevance check showed that not all items required major changes, but several stimuli needed stronger alignment with the everyday, occupational, and social experiences of Indonesian adults. Contextual revision therefore focused on improving the plausibility, familiarity, and authenticity of the situations rather than merely adding Indonesian surface features.

The FGD reviewed the 100 initial items in terms of content accuracy, stimulus clarity, and context relevance. The review ensured that contextual adaptation preserved the intended PIAAC competencies, that the stimuli and instructions were clear and unambiguous, and that the contexts were meaningful and authentic for Indonesian adults. Based on the FGD results, several items were revised and seven items were removed, reducing the item pool from 100 to 93 items. The summary of the FGD results is presented in Table 5.

Table 5. Results of the FGD stage

Aspect Reviewed	Findings from Self-Evaluation	Revision Decision
Content accuracy	Maintaining the cognitive demands specified in the PIAAC framework.	The team rechecked the mathematical procedures, corrected answer keys, clarified expected responses to maintain the cognitive demands and incorporating situations more relevant to Indonesian adults.
Stimulus clarity	Several stimuli were not yet clearly connected to the question or required clearer presentation.	Stimuli were revised, resized, or accompanied by more explicit instructions.
Context relevance	Most items used real-life contexts, but some contexts needed stronger alignment with adult situations.	Contexts were refined to represent better personal, work, and societal/community situations.

One-to-one

After the FGD stage, the revised items were tested through a one-to-one evaluation involving six prospective teachers. This stage examined the readability, clarity, and comprehensibility of the PIAAC items, as well as the initial usability of the computer-based testing (CBT) system. Participants were asked to complete the items and provide feedback on task difficulty, familiarity of the contexts, clarity of instructions, answer formats, and technical aspects of the system. The evaluation identified several issues, including unclear instructions regarding the expected answer format, rigid automated answer validation that rejected alternative but equivalent responses, visual information that was partially obstructed on the screen, and difficulties related to account management. These findings led to revisions in the item instructions, answer-validation mechanism, visual layout, and user-account procedures. An example of the rigid automated answer validation is presented in Figure 2, while Table 6 summarizes the readability results from the participants.

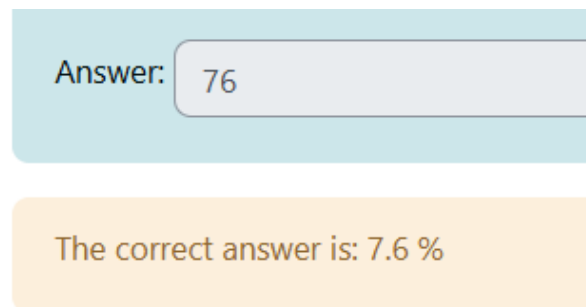


Figure 2. Rigid automated answer

Table 6. Recapitulatin of readability result

Aspect Evaluated	Summary of Findings	Response	Category	Follow-up Revision
Item difficulty	Most participants stated that the items were not too difficult and were at an easy-to-moderate level.	6/6 participants stated that the items were manageable.	Good	The items were retained, although some more difficult items were reviewed.
Relevance of item contexts	The contexts were considered familiar and close to participants' daily lives, such as pempek, Ampera, Jakabaring, health, business, and public services.	6/6 participants stated that the contexts were familiar.	Very Good	Local and contextual elements were retained.
Ease of using the CBT system	The CBT system was considered simple and relatively easy to use.	4/6 participants stated that the system was easy to use.	Good	The basic interface was retained.

Aspect Evaluated	Summary of Findings	Response	Category	Follow-up Revision
Clarity of instructions	Participants reported that some instructions, particularly regarding answer formats, remained unclear.	5/6 participants highlighted unclear answer-format instructions.	Needs revision	Clearer instructions were added before the test, especially for short-answer and numeracy items.
Flexibility of answer format	The system was too rigid in reading answers; differences in periods, commas, spaces, or writing formats caused correct answers to be marked incorrect.	5/6 participants reported this issue.	Needs revision	The system was revised to accept variations of correct answers.
Technical stability and display	Some technical issues were found, including errors after entering answers, images/text covered by icons, and the use of a single account by several participants.	3/6 participants reported technical issues.	Fair	Technical improvements were made to the display, answer validation, and account management.

Small Group

Following the one-to-one evaluation and subsequent revisions, the PIAAC items were tested in a small-group evaluation. This stage involved six lecturers from the Faculty of Teacher Training and Education. The purpose of the small-group evaluation was to examine the practicality of the developed items and to assess the readiness of the CBT system for broader implementation. While the one-to-one stage focused mainly on readability and comprehensibility, the small-group stage emphasized whether the items and the digital system were practical, usable, and sufficiently clear for adult participants. The results of the small-group evaluation indicated that the numeracy and adaptive problem-solving items were considered practical. Participants did not report substantial difficulties in understanding or completing these items. This suggests that the revisions made after the one-to-one stage had improved the clarity of the tasks and the usability of the item structure. The contexts, stimuli, and expected responses were generally considered appropriate for the intended adult users. [Table 7](#) presents a summary of the small-group test results.

Table 7. Summary of the small group evaluation results

Aspect Evaluated	Participants' Comments	Category	Follow-up Revision
Numeracy items	No major comments were given; the items were considered understandable and usable.	Practical	The numeracy items were retained.
Literacy items	One text had been provided, but the corresponding question had not yet been written.	Needs revision	The missing question was added, and the item's completeness was checked.
Adaptive problem-solving items	No major comments were given; the items were considered practical.	Practical	The adaptive problem-solving items were retained.
CBT system	The system was still somewhat rigid in recognizing variations of the correct answer.	Needs revision	The answer-validation system was revised to allow more flexible response formats.

3.1.3. Assessment Phase

The assessment phase was conducted through a field test of the developed PIAAC item prototype. This phase involved 30 Civic Education teachers from elementary, junior secondary, and senior secondary school levels. The purpose of the field test was to obtain an initial profile of participants' performance in completing PIAAC-based items across three competency domains: literacy, numeracy, and problem solving. The data collected in this phase consisted of participants' correct responses, which were then analyzed descriptively to examine their achievement in each domain.

Before presenting the participants' performance, the demographic characteristics of the teachers involved in the field test were first summarized (see [Table 8](#)). These characteristics included the number of participants, gender, and teaching experience. This information was important to provide a general description of the adult participants who engaged with the assessment environment.

Table 8. Demographic characteristics of participants in the assessment phase

No	Aspect	Category	Number	Percentage
1	Number of participants	Total	30	100%
2	Gender	Female	20	67%
3	Gender	Male	10	33%
4	Teaching experience	More than 10 years	0	0%
5	Teaching experience	5-10 years	0	0%
6	Teaching experience	Less than 5 years	30	100%

The field test data were analyzed by the number of correct answers across the three domains. Each domain consisted of 10 items, yielding a maximum score of 10 per domain and a maximum total score of 30. Descriptive analysis was conducted by calculating the mean score, percentage of achievement, median, minimum and maximum scores, standard deviation, and general interpretation of participants' achievement. The results are presented in [Table 9](#).

Table 9. Descriptive statistics of participants' performance in the assessment phase

Domain	N	Mean	Percentage	Median	Min-Max	SD	General Interpretation
Literacy	30	8.43	84.33%	8.5	4-10	1.22	High
Numeracy	30	7.43	74.33%	8	5-10	1.65	Moderate
Problem solving	30	8.03	80.33%	8	4-10	1.61	High

The results of the assessment phase indicate that, overall, the participants achieved relatively good performance across the three PIAAC competency domains. This finding suggests that the developed assessment environment, including the contextual PIAAC items and CBT system, engaged adult participants in tasks related to literacy, numeracy, and problem-solving.

Among the three domains, literacy had the highest mean score, at 8.43 out of 10 (84.33%). This indicates that the participants were relatively strong in accessing, understanding, and interpreting written information presented in the items. The problem-solving domain also showed a high level of achievement, with a mean score of 8.03 out of 10 (80.33%). This suggests that the participants were generally able to respond to dynamic situations, identify relevant constraints, and select appropriate strategies in contextual problem-solving tasks.

In contrast, numeracy had the lowest mean score, at 7.43 out of 10 (74.33%). Although this score still indicates a moderate level of achievement, it suggests that numeracy required the most attention. The numeracy items required participants to interpret quantitative information, use mathematical reasoning, compare numerical data, and make decisions based on calculations or representations such as tables, diagrams, and contextual data. The lower score in this domain may indicate that participants experienced greater difficulty when tasks involved numerical interpretation or mathematical decision-making in real-life contexts.

The distribution of scores also provides further insight into participants' performance. In the literacy domain, the scores ranged from 4 to 10, with a relatively low standard deviation of 1.22, indicating that most participants performed consistently well. In the problem-solving domain, the scores also ranged from 4 to 10, with a standard deviation of 1.61, showing slightly greater variation among participants. The numeracy domain had a score range of 5 to 10 and the highest standard deviation (1.65), suggesting that participants' numeracy performance varied more widely than their literacy performance.

Overall, the field test demonstrated that the developed PIAAC items were feasible for assessing adult competencies in contextual settings. The participants' high achievement in literacy and problem solving indicates that the items were accessible and relevant to their experiences as adult learners and professionals. However, the relatively low performance in

numeracy highlights the need for further refinement of numeracy items and additional support for adults in interpreting and using quantitative information in daily life situations. These findings provide an empirical basis for evaluating the effectiveness of the developed assessment environment and for improving the final version of the PIAAC item prototype.

Computer-Based Test (CBT) PIAAC

The development process also led to the establishment of the PIAAC Indonesia CBT (see Figure 3), built on the Moodle platform and customized to support the assessment of Indonesian adults' competencies in literacy, numeracy, and adaptive problem-solving. The CBT is accessible through the PIAAC Indonesia website (<https://piaac.id>) and serves as the primary digital platform for administering the developed assessment items. The platform was developed through the collaboration of the RKI Equity Program involving Universitas Sriwijaya (UNSRI), Universitas Negeri Padang (UNP), and Universitas PGRI Yogyakarta (UPY), with support from *Diktisaintek Berdampak*.

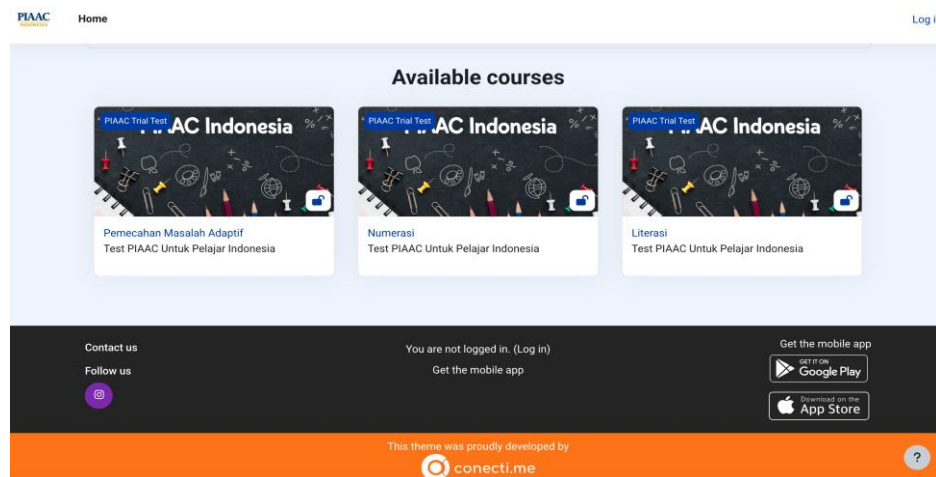


Figure 3. CBT PIAAC

Beyond functioning as a conventional computer-based testing platform, the PIAAC Indonesia CBT integrates assessment administration, learning resources, user management, automated scoring, and competency-based learning modules within a single environment. The homepage provides access to the three competency domains (literacy, numeracy, and adaptive problem solving), allowing users to navigate directly to the desired assessment module (see Figure 4). Participants can complete contextualized PIAAC tasks online, receive automated scoring, review their results, and access additional practice modules for independent learning. These integrated features position the platform not only as an assessment tool but also as a learning environment that supports continuous competency development.

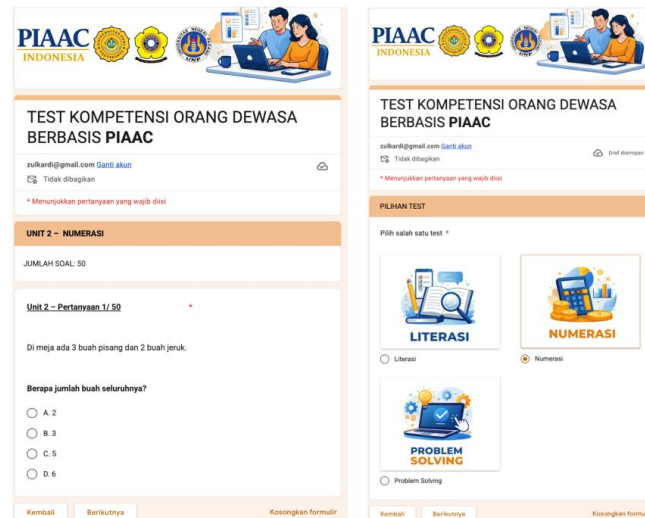


Figure 4. PIAAC test online

Another important outcome of the development is the establishment of an online training environment that extends PIAAC's functions beyond assessment. Although currently implemented primarily for competency testing, the CBT has been designed to accommodate future online courses and professional training in adult literacy, numeracy, and adaptive problem-solving (see Figure 5). Consequently, the platform represents an initial step toward the broader dissemination and sustainability of PIAAC-oriented learning resources in Indonesia.

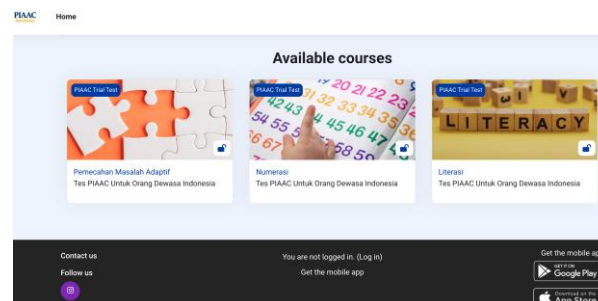


Figure 5. PIAAC online courses

Publication and Dissemination of Indonesian PIAAC Data through Books and Social Media

The PIAAC items developed in this study were disseminated through two primary publication formats: a book and PIAAC Indonesia's official social media channels. These two formats were used to make the items accessible to different user groups. The book provided a more systematic and comprehensive presentation of the items, whereas social media presented selected tasks in concise, visual, and readily accessible formats. Through this dual-format dissemination strategy, the PIAAC items were positioned not only as assessment materials but also as educational resources for introducing a broader audience to PIAAC-oriented literacy, numeracy, and adaptive problem-solving tasks.

The book publication compiled the developed items in a structured format, enabling users to examine the stimuli, questions, expected responses, and explanations in greater detail.

The book was organized around the major domains of the PIAAC framework, namely literacy, numeracy, and adaptive problem-solving. Each item was developed using information or situations that adults might encounter in personal, social, civic, or professional contexts. Accordingly, the book illustrated, through concrete examples, how adult competencies could be assessed using contextualized tasks rather than questions that merely measured the ability to recall school-based knowledge.

Publishing the items in book form also supported the documentation of key aspects of the item-development process. Unlike isolated test questions, the items presented in the book demonstrated the relationships among context, stimulus, cognitive processes, and the competencies being assessed. The items showed how written texts, images, tables, schedules, announcements, and other forms of information were transformed into tasks requiring interpretation, reasoning, and decision-making. The book could therefore serve as a reference for university lecturers, teachers, education professionals, researchers, and other stakeholders seeking to understand or develop assessment tasks aligned with the PIAAC framework.

Selected items were also adapted into visual content and published on PIAAC Indonesia's social media channels. As shown in [Figure 6](#), these social media posts featured familiar and contemporary situations, including football competitions, match results, and changes to public transportation routes. These contexts did not function merely as decorative backgrounds. Rather, they formed integral components of the tasks because users were required to identify relevant information, interpret the information presented, consider possible outcomes, or select an appropriate response.

For example, the football tournament item required users to interpret a competition bracket and predict the remaining stages of the tournament and the possible winner. This task involved reading a visual representation, recognizing the competition's structure, and making a reasoned prediction based on the available information. Another item used football match results as a stimulus for users to examine the presented information and provide an interpretation. Although the task appeared simple, it required users to distinguish between information that was directly observed and conclusions inferred from the stimulus. Similarly, the public transportation item required users to read an announcement concerning a route change and determine which route a passenger should take. This item integrated literacy and problem solving because the correct response depended on the ability to understand the announcement, compare the available routes, and apply the information to a specific travel situation.

These examples showed that social media could be used to communicate the characteristics of PIAAC tasks in a more immediate and familiar format for the public. The visual format also made the relationship between assessment tasks and everyday experiences more explicit by presenting the tasks through media commonly used by the public. At the same time, the posts retained the essential characteristics of PIAAC items, particularly the use of authentic information, real-life contexts, multiple forms of representation, and questions oriented toward decision-making. Therefore, the social media content was designed not only to promote the PIAAC Indonesia program but also to offer opportunities for the public to engage informally in literacy, numeracy, and problem-solving activities. [Figure 6](#) shows examples of the content published through PIAAC Indonesia's social media channels.



Figure 6. PIAAC Indonesia social media content

The book and social media formats served complementary functions. The book offered depth, systematic organization, and detailed documentation, making it suitable for professional learning, training, and further item development (see [Figure 7](#)). In contrast, social media emphasized accessibility, visual communication, and public engagement. Its more concise format enabled selected items to be introduced without requiring users to complete a full assessment instrument. This distinction indicated that the dissemination medium influenced how the items were presented, although the underlying competencies and contextual orientation remained consistent across formats.

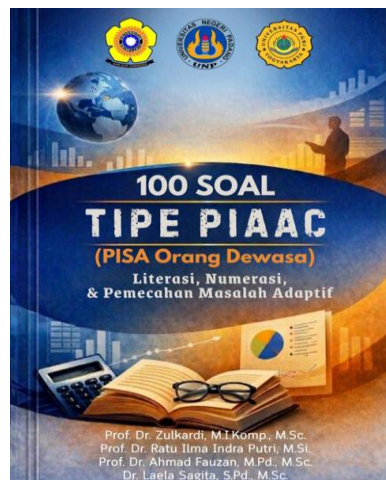


Figure 7. PIAAC problem book

The findings also suggested that the developed PIAAC items had potential functions beyond formal assessment. When published as a book and on social media, the items had the potential to increase public awareness of the competencies adults needed to process information and meet the demands of everyday life. Within the published materials, literacy was represented as the capacity to understand and evaluate information, numeracy as the capacity to interpret and apply quantitative relationships, and adaptive problem-solving as the capacity to select and adjust strategies in response to contextualized problems. This representation was consistent with the view that PIAAC assesses adults' capacity to apply knowledge and skills in functional situations rather than merely measuring mastery of academic knowledge in a context-free manner (Zulkardi et al., [2026](#)).

Nevertheless, publication alone did not provide evidence that the items improved users' competencies. At this stage, the results primarily documented the forms of the resulting products, the contextual characteristics of the items, and their potential for wider dissemination. Further evaluation is required to investigate how readers and social media users interpret the items, what difficulties they encounter, and whether repeated engagement with the published tasks can support the development of literacy, numeracy, or adaptive problem-solving competencies. Despite this limitation, the dual-format publication strategy represented an initial effort to connect PIAAC-oriented assessment with broader educational practices and public literacy initiatives in Indonesia.

PIAAC Dissemination through Courses and Workshops

In addition to digital dissemination through books and social media, the developed PIAAC assessment resources were introduced through face-to-face academic dissemination activities, including courses and workshops (see Figure 8). These activities aimed to familiarize lecturers, prospective teachers, teachers, and educational practitioners with the PIAAC framework and its contextualized assessment tasks. Rather than serving solely as dissemination events, the courses and workshops provided opportunities for participants to experience the developed assessment environment, discuss item characteristics, and understand the principles underlying authentic adult competency assessment.

The dissemination activities also support the sustainability of the developed assessment environment by encouraging broader adoption among educational institutions. Integrating workshops with the web-based platform enables participants to continue accessing the developed assessment resources beyond the training sessions, thereby extending the project's educational impact.



Figure 8. PIAAC onsite dissemination

3.2. Discussion

The findings demonstrate that contextualizing the PIAAC framework requires more than merely translating international assessment items into Indonesian. The central design challenge was to preserve the competencies, cognitive processes, and task characteristics underlying PIAAC while transforming the stimuli and situations into contexts recognizable to Indonesian adults. The resulting items incorporated personal, occupational, and societal

situations and required participants to access information, interpret representations, evaluate alternatives, use quantitative reasoning, and adjust strategies. This supports the fundamental orientation of PIAAC, in which adult competence is understood as the capacity to mobilize knowledge and skills in functional situations rather than merely reproduce school-based knowledge (Kirsch et al., 2020; OECD, 2013). Therefore, contextualization in this study served as an essential component of task design rather than a superficial replacement for names, places, or illustrations.

In this study, contextualization was therefore conducted at the level of the stimulus and situational setting rather than at the level of the competency being assessed. For example, a numeracy task could replace an unfamiliar situation with information related to household expenses, transportation, or local public services while retaining the requirement to interpret quantitative information and make a reasoned decision. Similarly, literacy stimuli could be adapted to texts commonly encountered by Indonesian prospective teachers while preserving the intended processes of accessing, understanding, and evaluating information. In adaptive problem solving, the scenario could be localized while maintaining the need to identify constraints, evaluate alternatives, and adjust strategies. Thus, contextualization involved changing the surface features of the task without altering its underlying competency and cognitive demand.

This distinction is particularly important in adult assessment because authentic tasks should represent situations in which adults are likely to use these competencies in their personal, occupational, and societal lives. Contextual relevance therefore does not simply make an item more familiar; it helps situate the targeted competency within realistic adult practices while maintaining the intended cognitive challenge.

The development of the items also revealed that alignment among competency domains, cognitive processes, contexts, stimuli, response formats, and expected answers must be explicitly documented. The use of a standardized item structure helped the researchers determine the dominant competency measured by each task, particularly when the characteristics of literacy, numeracy, and problem solving overlapped. This was important because authentic tasks commonly require more than one form of information-processing competence. A numeracy item, for example, may also require substantial reading, while a problem-solving task may involve interpreting numerical data. Rather than eliminating this overlap, the design process identified the dominant construct and ensured that the required cognitive process corresponded with the intended domain. This finding is consistent with research showing that PIAAC tasks comprise different forms and levels of cognitive engagement and should therefore be analysed in terms of what respondents are required to do with the information presented (Boeren & Íñiguez-Berrozpe, 2022).

The iterative prototyping process was essential for translating the conceptual framework into a usable assessment environment. The focus group discussion identified conceptual and structural issues, including inaccurate answer keys, incomplete item information, overly general cognitive-process labels, and inconsistencies between item contexts and competency domains. The one-to-one evaluation subsequently revealed difficulties that were less visible during internal review, particularly unclear answer-format instructions, rigid automatic answer validation, obstructed visual information, and account-

management problems. The small-group evaluation confirmed that most literacy, numeracy, and adaptive problem-solving tasks were understandable and practical after revision, although further improvements were still required in item completeness and system flexibility. These results illustrate the value of formative evaluation in Educational Design Research because product quality emerged through successive cycles of implementation, feedback, and revision rather than through a single validation procedure. This corresponds with the view of McKenney and Reeves (2025) that educational interventions should evolve through continuous interaction between theoretical intentions and evidence from practical use.

The usability findings further indicate that the technical design of a computer-based assessment can affect the validity of the evidence collected. In the one-to-one evaluation, correct responses could be classified as incorrect because of differences in punctuation, spacing, decimal notation, or writing format. Participants also encountered unclear instructions and visual elements that were obscured by interface icons. These problems were not directly related to literacy, numeracy, or problem-solving ability, yet they could interfere with performance and produce construct-irrelevant difficulty. Revising the answer-validation system to accommodate equivalent response forms and improving the interface were therefore not merely technical adjustments; they were necessary for ensuring that the assessment more accurately represented the competencies it intended to measure. This finding reinforces the argument that assessment in technology-rich environments must consider navigation, response mechanisms, information presentation, and user interaction as integral parts of task design.

The familiarity of the contexts also contributed to the accessibility of the developed items. During the one-to-one evaluation, participants regarded contexts involving pempek, the Ampera Bridge, Jakabaring, health, business, and public services as closely related to their everyday experiences. Familiar contexts may help respondents understand the situation without requiring extensive background knowledge unrelated to the intended competency. Nevertheless, contextual familiarity should not be interpreted as making an item automatically easy. Authentic contexts still require participants to distinguish relevant from irrelevant information, understand constraints, select strategies, and justify decisions. Thus, the function of contextualization is not to reduce cognitive demand but to make that demand meaningful and representative of how adult competencies are used outside formal educational settings. This interpretation is consistent with Ashford-Rowe et al. (2014), who emphasize realism, contextual application, and meaningful performance as central features of authentic assessment.

The field-test results provide initial evidence that the developed assessment environment was feasible for adult participants. Literacy obtained the highest mean score, followed by problem solving and numeracy. Participants' relatively strong literacy performance suggests that they were generally able to access, understand, and interpret written information in contextualized stimuli. Their high performance in problem solving indicates an ability to recognize constraints, evaluate available options, and choose appropriate responses in situational tasks. Numeracy, although still categorized as moderate, had the lowest mean and the greatest score variation. The numeracy tasks required participants to interpret tables, diagrams, financial information, measurements, and quantitative relationships and to use these representations in decision-making. Accordingly, the lower numeracy performance may

indicate that applying mathematical information in contextual situations was more demanding for participants than interpreting written information.

The field-test results showed differences in performance across the three competency domains. Literacy obtained the highest mean score (8.43), followed by problem solving (8.03) and numeracy (7.43). Numeracy also showed the greatest score variation among the three domains. These results indicate that, in this field test, participants achieved lower scores on the numeracy items than on the literacy and problem-solving items. One possible interpretation of the lower numeracy performance is related to the nature of contextual numeracy tasks. Such tasks may require respondents to coordinate several processes, including identifying relevant quantitative information, interpreting representations, selecting appropriate procedures, and using the results to make decisions. This interpretation is theoretically consistent with the literature on contextual numeracy. However, it cannot be confirmed from the present data. The study did not collect data that could identify the specific reasons for participants' difficulty. Therefore, other explanations, such as stimulus complexity, linguistic demands, item difficulty, or features of the digital response format, also remain possible. This finding has an important practical implication: future refinement should examine whether item difficulty stems from the intended mathematical reasoning, the complexity of the stimulus, linguistic demands, or the digital response format. Larger-scale item analysis is consequently required before the tasks can be used for high-stakes measurement.

The field test involved 30 Civic Education teachers, and each participant completed ten items in each competency domain. Literacy achieved a mean of 8.43, problem solving 8.03, and numeracy 7.43 out of ten; numeracy also had the highest standard deviation. Therefore, these results should be interpreted as an initial performance profile and evidence of product feasibility, rather than definitive evidence of population-level adult competence or psychometric effectiveness.

Another contribution of the study lies in integrating the CBT system with an item book and social media content. These media served different but complementary purposes. The CBT system enabled participants to interact directly with contextualized tasks and provided structured performance data. The item book documented the relationship among competencies, contexts, cognitive processes, stimuli, answer formats, and expected responses, making it potentially useful for professional learning and future item development. Social media transformed selected items into concise and visually accessible materials that could introduce PIAAC-oriented competencies to a broader audience. This multimodal strategy extends the role of the developed items beyond formal testing by presenting literacy, numeracy, and adaptive problem solving as competencies embedded in everyday decision-making.

However, the dissemination findings should be interpreted cautiously. Publishing the materials in book and social media formats demonstrates accessibility and potential educational use, but it does not yet establish that engagement with these materials improves adult competencies. The article itself recognizes that further research is required to examine how users interpret the published tasks, what difficulties they encounter, and whether repeated interaction produces measurable learning outcomes. The educational contribution of the dissemination media is therefore currently situated in competency awareness, access to

examples, and support for professional discussion rather than in demonstrated competency growth.

Taken together, the study generates four interrelated design principles for contextualized PIAAC assessment environments. First, the theoretical constructs and cognitive processes of the international framework should be retained, while the situations and stimuli are reconstructed to center on locally meaningful adult practices. Second, every item should explicitly align its competency domain, cognitive process, context, stimulus, response format, expected response, and estimated difficulty. Third, formative evaluation should involve both conceptual review and direct user interaction because content- and interface-related problems emerge at different stages of development. Fourth, digital assessment should be supported by flexible response mechanisms and complementary dissemination formats without assuming that dissemination alone produces learning. These principles constitute the study's main design contribution and may inform similar efforts to adapt international competency frameworks in other cultural and educational contexts.

4. CONCLUSION

This study successfully produced contextualized PIAAC-type assessment resources that were valid and practical, and that demonstrated potential effects on adults' literacy, numeracy, and adaptive problem-solving competencies, particularly among lecturers, prospective teachers, and teachers in Palembang. The development process began with an initial pool of 100 items, from which 93 items were retained as the final contextualized assessment items, comprising 32 literacy, 31 numeracy, and 30 problem-solving items. The main outputs comprised a book containing the 93 final PIAAC-type items, a web-based assessment system accessible through piaac.id and Google Forms, social media dissemination through Instagram, and direct dissemination activities through courses and workshops. These products provide an integrated environment for assessing, introducing, and strengthening adult competencies through authentic personal, professional, and societal contexts.

Further development should focus on transforming the web-based assessment system from a testing platform into an online training environment that explains the purpose, relevance, and application of PIAAC competencies in everyday and workplace situations. Dissemination should also be expanded through the CBT and additional social media platforms, including X, to reach a wider population of learners and workers aged 16–65. In addition, PIAAC-oriented content should be integrated as a regular component of the PMRI–PISA Summer Course and relevant university courses to support the sustainability and broader educational use of the developed resources.

Acknowledgments

The authors would like to extend their sincere gratitude to the *Asosiasi Lembaga Penelitian dan Pengabdian Masyarakat Perguruan Tinggi Negeri Badan Hukum* for the financial support provided through the “Pendanaan Program Riset Kolaborasi Indonesia Equity Skema A1, A2, B1, B2 Dana Abadi Pendidikan Tinggi Lembaga Pengelola Dana Pendidikan Tahun 2025-2026”, under the contract 001/ALP/HK/XI/2025.

Declarations

- Author Contribution** : Z: Conceptualization, Visualization, Writing - original draft, and Writing - review & editing; RIIP: Methodology, and Writing - review & editing; AF: Methodology, and Writing - review & editing; LS: Methodology, and Writing - review & editing; RSU: Formal analysis, Methodology, and Writing - review & editing; VA: Formal analysis, Methodology, and Writing - review & editing; YS: Formal analysis, Methodology, and Writing - review & editing; DAA: Formal analysis, Methodology, and Writing - review & editing; DAP: Formal analysis, Methodology, and Writing - review & editing; NF: Formal analysis, Methodology, and Writing - review & editing; NS: Formal analysis, Methodology, and Writing - review & editing; RA: Formal analysis, Methodology, and Writing - review & editing.
- Funding Statement** : This research was funded by the *Asosiasi Lembaga Penelitian dan Pengabdian Masyarakat Perguruan Tinggi Negeri Badan Hukum* through the “Pendanaan Program Riset Kolaborasi Indonesia Equity Skema A1, A2, B1, B2 Dana Abadi Pendidikan Tinggi Lembaga Pengelola Dana Pendidikan Tahun 2025-2026”, under the contract 001/ALP/HK/XI/2025.
- Conflict of Interest** : The authors declare no conflict of interest.
- Additional Information** : Additional information is available for this paper.

REFERENCES

- Abbasi, H. (2025). Educator perspectives on the transformative power of authentic assessments on student engagement, empowerment, and inclusivity. *Educational Practice and Theory*, 47(1), 41–54. <https://doi.org/10.7459/ept/470104>
- Ashford-Rowe, K., Herrington, J., & Brown, C. (2014). Establishing the critical elements that determine authentic assessment. *Assessment & Evaluation in Higher Education*, 39(2), 205–222. <https://doi.org/10.1080/02602938.2013.819566>
- Boeren, E., & Íñiguez-Berrozpe, T. (2022). Unpacking PIAAC’s cognitive skills measurements through engagement with Bloom’s taxonomy. *Studies in Educational Evaluation*, 73, 101151. <https://doi.org/10.1016/j.stueduc.2022.101151>
- Carnevale, A. P., & Smith, N. (2013). Workplace basics: The skills employees need and employers want. *Human Resource Development International*, 16(5), 491–501. <https://doi.org/10.1080/13678868.2013.821267>
- Chen, Y., & Nunes, B. P. (2025). Quantifying assessment difficulty in GenAI-enhanced education: Integrating reinterpreted RBT and SOLO taxonomy. In *2025 IEEE International Conference on Advanced Learning Technologies (ICALT)*, (pp. 382–386). <https://doi.org/10.1109/icalt64023.2025.00118>
- Cosser, D. M. (2018). Differential pathways of South African students through higher education. *International Journal of Educational Development*, 59, 100–109. <https://doi.org/10.1016/j.ijedudev.2017.10.003>

- De Wever, B., Hämäläinen, R., Nissinen, K., Mannonen, J., & Van Nieuwenhove, L. (2023). Teachers' problem-solving skills in technology-rich environments: A call for workplace learning and opportunities to develop professionally. *Studies in Continuing Education*, 45(1), 86–112. <https://doi.org/10.1080/0158037x.2021.2003769>
- Díaz-García, V., Montero-Navarro, A., Rodríguez-Sánchez, J.-L., & Gallego-Losada, R. (2022). Digitalization and digital transformation in higher education: A bibliometric analysis. *Frontiers in psychology*, 13, 1081595. <https://doi.org/10.3389/fpsyg.2022.1081595>
- Elwy, A. R., Wasan, A. D., Gillman, A. G., Johnston, K. L., Dodds, N., McFarland, C., & Greco, C. M. (2020). Using formative evaluation methods to improve clinical implementation efforts: Description and an example. *Psychiatry Research*, 283, 112532. <https://doi.org/10.1016/j.psychres.2019.112532>
- García-Peñalvo, F. J. (2022). Desarrollo de estados de la cuestión robustos: Revisiones Sistemáticas de Literatura [Developing robust state-of-the-art reviews: Systematic literature reviews]. *Education in the Knowledge Society (EKS)*, 23, e28600. <https://doi.org/10.14201/eks.28600>
- Goldhammer, F., Hahnel, C., & Kroehne, U. (2020). Analysing log file data from PIAAC. In D. B. Maehler & B. Rammstedt (Eds.), *Large-scale cognitive assessment: Analyzing PIAAC data* (pp. 239–269). Springer. https://doi.org/10.1007/978-3-030-47515-4_10
- González-Such, J., Sancho-Álvarez, C., & Sánchez-Delgado, P. (2016). Cuestionarios de contexto PISA: Un estudio sobre los indicadores de evaluación [PISA context questionnaires: A study on assessment indicators]. *RELIEVE - Revista Electrónica de Investigación y Evaluación Educativa*, 22(1), 1–26. <https://doi.org/10.7203/relieve.22.1.8274>
- Gorges, J., Koch, T., Maehler, D. B., & Offerhaus, J. (2017). Same but different? Measurement invariance of the PIAAC motivation-to-learn scale across key socio-demographic groups. *Large-scale Assessments in Education*, 5(1), 13. <https://doi.org/10.1186/s40536-017-0047-5>
- He, Q., & von Davier, M. (2015). Identifying feature sequences from process data in problem-solving items with N-grams. In L. van der Ark, D. Bolt, W. C. Wang, J. Douglas, & S. M. Chow (Eds.), *Quantitative Psychology Research* (Vol. 140, pp. 173–190). Springer. https://doi.org/10.1007/978-3-319-19977-1_13
- Jakubowski, M., & Pokropek, A. (2019). Piaactools: A program for data analysis with PIAAC data. *The Stata Journal: Promoting communications on statistics and Stata*, 19(1), 112–128. <https://doi.org/10.1177/1536867x19830909>
- Kim, S. (2018). Literacy skills gaps: A cross-level analysis on international and intergenerational variations. *International Review of Education*, 64(1), 85–110. <https://doi.org/10.1007/s11159-018-9703-4>
- Kirsch, I., & Lennon, M. L. (2017). PIAAC: A new design for a new era. *Large-scale Assessments in Education*, 5(1), 11. <https://doi.org/10.1186/s40536-017-0046-6>
- Kirsch, I., Yamamoto, K., & Khorrarnadel, L. (2020). Design and key features of the PIAAC survey of adults. In D. B. Maehler & B. Rammstedt (Eds.), *Large-scale cognitive assessment* (pp. 7–26). Springer. https://doi.org/10.1007/978-3-030-47515-4_2

- Leighton, J. P., Heffernan, C., Cor, M. K., Gokierto, R. J., & Cui, Y. (2011). An experimental test of student verbal reports and teacher evaluations as a source of validity evidence for test development. *Applied Measurement in Education*, 24(4), 324–348. <https://doi.org/10.1080/08957347.2011.607061>
- Liao, D., He, Q., & Jiao, H. (2019). Mapping background variables with sequential patterns in problem-solving environments: An investigation of United States adults' employment status in PIAAC. *Frontiers in psychology*, 10, 00646. <https://doi.org/10.3389/fpsyg.2019.00646>
- Lopes, T., Scully-Russ, E., Zarestky, J., & Collins, J. C. (2020). The effects of social characteristics of jobs on the cognitive skills of adults in the United States: A PIAAC analysis. *Adult education quarterly*, 70(2), 140–174. <https://doi.org/10.1177/0741713619884567>
- Makeham, S., & Lee, C. (2012). Making the aural presentation of examination papers student friendly: An alternative to a reader in examinations. *Assessment & Evaluation in Higher Education*, 37(2), 237–243. <https://doi.org/10.1080/02602938.2010.527915>
- McKenney, S., & Reeves, T. C. (2025). Educational design research for relevant & robust scholarship. *Journal of Computing in Higher Education*, 37(2), 614–638. <https://doi.org/10.1007/s12528-025-09456-2>
- Mokkink, L. B., Herbelet, S., Tuinman, P. R., & Terwee, C. B. (2025). Content validity: Judging the relevance, comprehensiveness, and comprehensibility of an outcome measurement instrument – a COSMIN perspective. *Journal of clinical epidemiology*, 185, 111879. <https://doi.org/10.1016/j.jclinepi.2025.111879>
- OECD. (2013). *OECD skills outlook 2013: First results from the survey of adult skills*. OECD Publishing. <https://doi.org/10.1787/9789264204256-en>
- OECD. (2016). *Skills matter: Further results from the survey of adult skills*. OECD Publishing. <https://doi.org/10.1787/9789264258051-en>
- OECD. (2021). Investing in competences and skills and reforming the labour market to create better jobs in Indonesia. *OECD Economics Department Working Papers*, No. 1670, OECD Publishing. <https://doi.org/10.1787/fd54e6be-en>
- OECD. (2024). *Do adults have the skills they need to thrive in a changing world?: Survey of adult skills 2023*. OECD Publishing. <https://doi.org/10.1787/b263dc5d-en>
- Oroh, E. Z., Ali, M. I., Pelenkahu, N., Usman, H., & Rorintulus, O. A. (2025). Authentic assessment in higher education to increase critical thinking and develop metacognitive awareness. *Studies in English Language and Education*, 12(2), 827–844.
- Perry, K. H., Shaw, D. M., & Saberimoghaddam, S. (2020). Literacy practices and the programme for the international assessment of adult competencies (PIAAC): A conceptual critique. *International Review of Education*, 66(1), 9–28. <https://doi.org/10.1007/s11159-019-09819-9>
- Plomp, T. (2013). Educational design research: An introduction. In T. Plomp & N. Nieveen (Eds.), *Educational design research* (pp. 10–51). Netherlands Institute for Curriculum Development (SLO).
- Rammstedt, B., Perry, A., & Maehler, D. B. (2015). Zentrale ergebnisse von PIAAC aus deutscher perspektive [Key findings from PIAAC from a German perspective]. *Zeitschrift für pädagogik*, 61(2), 162–191.

- Ruiz-Primo, M. A., & Li, M. (2016). Ítems contextualizados de PISA Ciencias: La conexión entre las demandas cognitivas y las características de contexto de los ítems [Contextualized PISA science items: The connection between cognitive demands and item context characteristics]. *RELIEVE - Revista Electrónica de Investigación y Evaluación Educativa*, 22(1), 13. <https://doi.org/10.7203/relieve.22.1.8280>
- Sagita, L., Utami, N. W., Dwipa, N. M. S., Zulkardi, Z., Putri, R. I. I., & Fauzan, A. (2026). Designing a contextual numeracy assessment for adult learners: Evidence from readability testing based on the PIAAC framework. *Mathematics Education Journal*, 20(3), 671–692. <https://doi.org/10.22342/mej.v20i3.pp671-692>
- Schleicher, A. (2008). Piaac: A new strategy for assessing adult competencies. *International Review of Education*, 54(5-6), 627–650. <https://doi.org/10.1007/s11159-008-9105-0>
- Turk, N. (2021). Metodologija priprave sistematičnih preglednih člankov [Methodology for preparing systematic review articles]. *Slovenian Medical Journal*, 90(7-8), 432–442. <https://doi.org/10.6016/ZdravVestn.3138>
- van Vugt, L., Levels, M., & van der Velden, R. (2024). The low skills trap: The failure of education and social policies in preventing low-literate young people from being long-term NEET. *Journal of Youth Studies*, 27(2), 217–251. <https://doi.org/10.1080/13676261.2022.2118036>
- Venancio, S. I., Bortoli, M. C., Frias, P. G., Giugliani, E. R. J., Alves, C. R. L., & Santos, M. O. (2020). Development and validation of an instrument for monitoring child development indicators. *Jornal de Pediatria*, 96(6), 778–789. <https://doi.org/10.1016/j.jped.2019.10.008>
- Villarroel, V., Bloxham, S., Bruna, D., Bruna, C., & Herrera-Seda, C. (2018). Authentic assessment: Creating a blueprint for course design. *Assessment & Evaluation in Higher Education*, 43(5), 840–854. <https://doi.org/10.1080/02602938.2017.1412396>
- Vörös, Z., & Rouet, J.-F. (2016). Laypersons' digital problem solving: Relationships between strategy and performance in a large-scale international survey. *Computers in human Behavior*, 64, 108–116. <https://doi.org/10.1016/j.chb.2016.06.018>
- Yamashita, T., Bardo, A. R., Cummins, P. A., Millar, R. J., Sahoo, S., & Liu, D. (2019). The roles of education, literacy, and numeracy in need for health information during the second half of adulthood: A moderated mediation analysis. *Journal of Health Communication*, 24(3), 271–283. <https://doi.org/10.1080/10810730.2019.1601303>
- Zulkardi, Z., Putri, R. I. I., Fauzan, A., Sagita, L., Alfajri, R., Silvhiany, S., Adelia, V., Anggeraini, Y., Putri, D. A., Utami, N. W., Dwipa, N. M. S., Utari, R. S., Sukma, Y., Azka, D. A., Ismail, R. N., Sari, N., Asra, A., & Fitriani, N. (2026). Assessment frameworks utilized by PIAAC for evaluating adult problem-solving skills. *Mathematics Education Journal*, 20(2), 423–448. <https://doi.org/10.22342/mej.v20i2.pp423-448>