Formatif: Jurnal Ilmiah Pendidikan MIPA, September 2025, 15 (2), 1-18
https://doi.org/10.30998/1r4awm79
p-ISSN: 2088-351X e-ISSN: 2502-5457

Available online at https://newjournal.lppmunindra.ac.id/index.php/formatif/index

Instrument Development: Problem Solving, Independence, and Communication Skills, integrated with Bakpia Making in the Material on Quantity and Measurement

Nisya Indriyanti Irawan[1*)], Ayu Mar'ati B.N[2], Edi Istiyono[3]
[1,2,3]University Negeri of Yogyakarta, Yogyakarta, Indonesia

Abstract
Assessing 21st-century skills such as problem solving, independent learning, and communication poses a significant challenge, as traditional assessments often fail to measure them holistically. This study aims to fill this gap by developing and validating a set of integrated assessment instruments—cognitive, affective, and psychomotor—contextualized with the local wisdom of bakpia making. The goal is to produce valid and reliable instruments that meet modern psychometric standards and can be applied to high school students. This study uses a quantitative development method, adapting the Wilson, Oriondo and Antonio model. The process includes three main stages: instrument design (developing a framework and items based on relevant theory), expert validation, and field testing on 10th-grade students at a public high school in Yogyakarta. The data were analyzed using the Rasch Model (Model 1-PL for cognitive; Graded Response Model for affective and psychomotor) to evaluate item fit, reliability, difficulty index, and item discrimination. The results showed that all three instruments had consistently strong psychometric properties. The cognitive instrument (problem solving) was valid, with item INFIT MNSQ values ranging from 0.80 to 1.27. This instrument was also highly reliable (Cronbach's alpha = 0.78) and had a good difficulty level distribution. The affective instrument (learning independence) was also proven to be valid (INFIT MNSQ = 0.86–1.18) and highly reliable (Cronbach's alpha = 0.90). Similarly, the psychomotor instrument (communication) is valid with an INFIT MNSQ value of 0.82–1.19 and shows exceptional reliability (Cronbach's alpha = 0.95). The positive item discrimination values on all instruments further reinforce their quality. In conclusion, this study successfully developed a valid, reliable, and contextually relevant integrated assessment instrument to measure 21st-century skills. These findings provide a practical model for educators to implement culture-based holistic assessments, which offer a comprehensive evaluation of student competencies beyond factual knowledge.

Keywords: Instrument, Problem Solving, Independence, Communication Skills, Quantity and Measurement

(*) Corresponding Author: nisyaindriyanti.2024@student.uny.ac.id

How to Cite: Irawan et al. (2025). Instrument Development: Problem Solving, Independence, and Communication Skills, integrated with Bakpia Making in the Material on Quantity and Measurement. *Formatif: Jurnal Ilmiah Pendidikan MIPA,* 15 (2): 1-18. https://doi.org/10.30998/1r4awm79

INTRODUCTION

21st-century education faces unprecedented challenges and opportunities. In an era marked by technological disruption, information abundance, and the complexity of global issues, every individual must be able to adapt, innovate, and collaborate (Griffin, 2012; Partnership for 21st Century Skills., 2019; Unik Hanifah Salsabila et al., 2020). Therefore, the purpose of education has evolved, no longer focusing solely on mastering academic material, but rather on developing individuals who possess 21st-century skills (such as critical thinking, collaboration, communication, and creativity) to face complex global challenges(Partnership for 21st Century Skills., 2019; Realitawati, 2024; Rifa Hanifa Mardhiyah et al., 2021). Among these essential skills, problem-solving and communication are important foundations.

The capacity to analyze complex situations, develop innovative solutions, and resolve obstacles is an essential cognitive skill in every aspect of life (Jonassen, 2000;

Polya, 1973). Ideally, problem-solving skills should be assessed at each step of the solution process. However, in schools, the dominant instrument is direct formula questions without analysis, which are ineffective for measuring these skills (Angriani et al., 2018). In addition, teachers rarely give problem-solving questions (Fitrianty et al., 2022). The teacher's approach to problem-solving is still evident in the classroom. After the demonstration, students practice following the steps shown (Prabawanto, 2019). Accordingly, communication skills—both verbal and written—are crucial for interacting, collaborating, and communicating ideas effectively in an increasingly connected and diverse society. (Hymes, 1972; Spenard, 2014). However, current assessment practices are sometimes still dominated by factual knowledge measurement, so they are not yet fully capable of capturing the depth of thinking processes in problem solving or measuring practical communication skills authentically. (Nitko & Brookhart, 2011).

Responding to these global demands, the Indonesian government introduced the Merdeka Curriculum as an educational transformation initiative focused on developing students' potential holistically. One of the characteristics of the Merdeka curriculum is the existence of a project to strengthen the Pancasila profile (Asiati & Hasanah, 2022; Sukmadana, 2024). The Pancasila student profile is a desired graduate profile that aims to demonstrate the character and competencies that students are expected to achieve. In addition, the Pancasila student profile also serves to strengthen students with the noble values of Pancasila (Mery et al., 2022; Rusnaini et al., 2021). The Pancasila student profile has six main elements, such as: 1) Being devoted to God Almighty and having noble character, 2) Being globally diverse, 3) Being cooperative, 4) Being independently minded, 5) Being critical thinkers, 6) Being creative(Amir et al., 2022; Asiati & Hasanah, 2022; Septiani, 2022). Independence in learning is an essential aspect of learning (Astuti & Fitriani, 2021; Permata et al., 2022). Independent learning is a serious issue that is one of the key factors in learning success, involving motivation, cognition, behavior, and student performance (You & Kang, 2014). The issue in the field related to the measurement of learning independence is the absence of standard instruments to assess student learning independence (Audhiha et al., 2022). Considering that the assessment of student learning outcomes covers three domains, including attitude assessment, content knowledge assessment, and skill assessment (Nugroho & Mawardi, 2021). The misuse of inaccurate testing instruments will result in inaccurate measurement results (Kurniawan et al., 2018).

Recognizing the urgency of developing cognitive (problem-solving), psychomotor (communication), and affective (independent learning) skills, an innovative, integrated assessment approach relevant to the learners' context is required. One promise approach is the integration of local wisdom as a medium for learning and assessment. This integration not only makes the material more meaningful and relevant but also contributes to the preservation of the nation's cultural heritage (Lesmana & Nurussaniah, 2022; Sae et al., 2021). Research by Hasyim et al. (2022) successfully developed a valid and reliable science literacy-based physics instrument using the context of Makassar's traditional cuisine. This study shows that an authentic cultural context can be a strong foundation for cognitive assessment. Furthermore, the feasibility of this approach was reaffirmed by research from Murti & Sunarti (2021) and Maulida & Sunarti (2022). Murti & Sunarti (2021) developed a science literacy instrument based on the local wisdom of Trenggalek, focusing on physical phenomena in traditional ceremonies such as Larung Sembonyo, proving that the instrument is theoretically and empirically feasible. Similarly, Maulida & Sunarti (2022) successfully developed a similar instrument by adapting local wisdom from Lamongan Regency, such as the games “kekehan” and “deligrang.” Both studies emphasize that culture-based instruments can be effective in measuring students' scientific skills. Not limited to science literacy, this approach has also been successfully applied to other cognitive skills. Nurkholifah et al. (2018) developed a critical thinking instrument that uses the context of traditional gong musical instruments for physics material. Meanwhile, Alfika et al. (2018) especifically focused on developing problem-solving instruments based on local wisdom. These studies provide a strong foundation that local wisdom can be an effective basis for measuring various higher-order cognitive skills.

In this case, bakpia making can be part of the learning process (Nisa & Wilujeng, 2021). In terms of cognition, making bakpia requires problem-solving skills related to quantities and measurements. For example, problems may arise when determining the right measuring tool to measure the diameter of bakpia or ensuring consistent use of the SI system to avoid errors in recipes and production. The aspect of independent learning is also greatly stimulated. In the context of learning, students will develop independence when they are given teaching materials or learning based on local wisdom (Nurhikmayati & Sunendar, 2020). Furthermore, communication skills will be honed through the integration of local wisdom in bakpia making (Nisa & Wilujeng, 2020), For instance, when practicing the production and measurement of bakpia. They are encouraged to express their opinions, ask questions, or share their observations regarding the process and measurements. It is at this moment that their communication skills can be measured and reflected upon, to what extent they can interact and collaborate effectively.

The authentic context of bakpia production provides a strong foundation for developing an assessment instrument capable of capturing all three dimensions simultaneously. This instrument can create assessment scenarios that are not only constructively valid but also relevant and engaging for students, while providing a holistic picture of their competencies. Therefore, this article aims to develop cognitive assessment instruments based on problem-solving, affective instruments based on independent learning, and psychomotor instruments based on communication, all of which are integrated within the context of local wisdom in bakpia production.

METHODS

This research is quantitative development research that focuses on instrument development. The model used for the development of this instrument is a modification of the Wilson Model and the Oriondo and Antonio Model (Istiyono et al., 2014) . This research was conducted in April and May 2025. The initial development phase of the test, including preparation, validation, and test design, was carried out from April to May 2025. Subsequently, the pilot test was conducted in May 2025. The entire study was conducted at high schools in the Yogyakarta region. The study subjects were selected from all tenth-grade students at three public high schools located in Yogyakarta. The total number of study subjects was 161 students.

This research implements the development of test instruments through modifications to the Wilson Model and the Oriondo and Antonio Model, which are broadly divided into three main stages: test design, test piloting, and test assembly. The test design stage begins with determining the objectives, competencies, and content to be assessed, followed by the development of a test blueprint and the creation of questions based on the skills to be measured, including cognitive problem-solving skills, affective learning independence, and psychomotor communication skills. The questions are then validated, revised, assembled, and supplemented with scoring guidelines. After that, the test piloting stage is conducted by selecting test subjects at high schools, conducting the testing, and analyzing the results. Finally, after all these processes, the final test is assembled to produce an instrument ready for use. This development process can be seen in the flowchart developed by Istiyono et al. (2014) in the following figure.

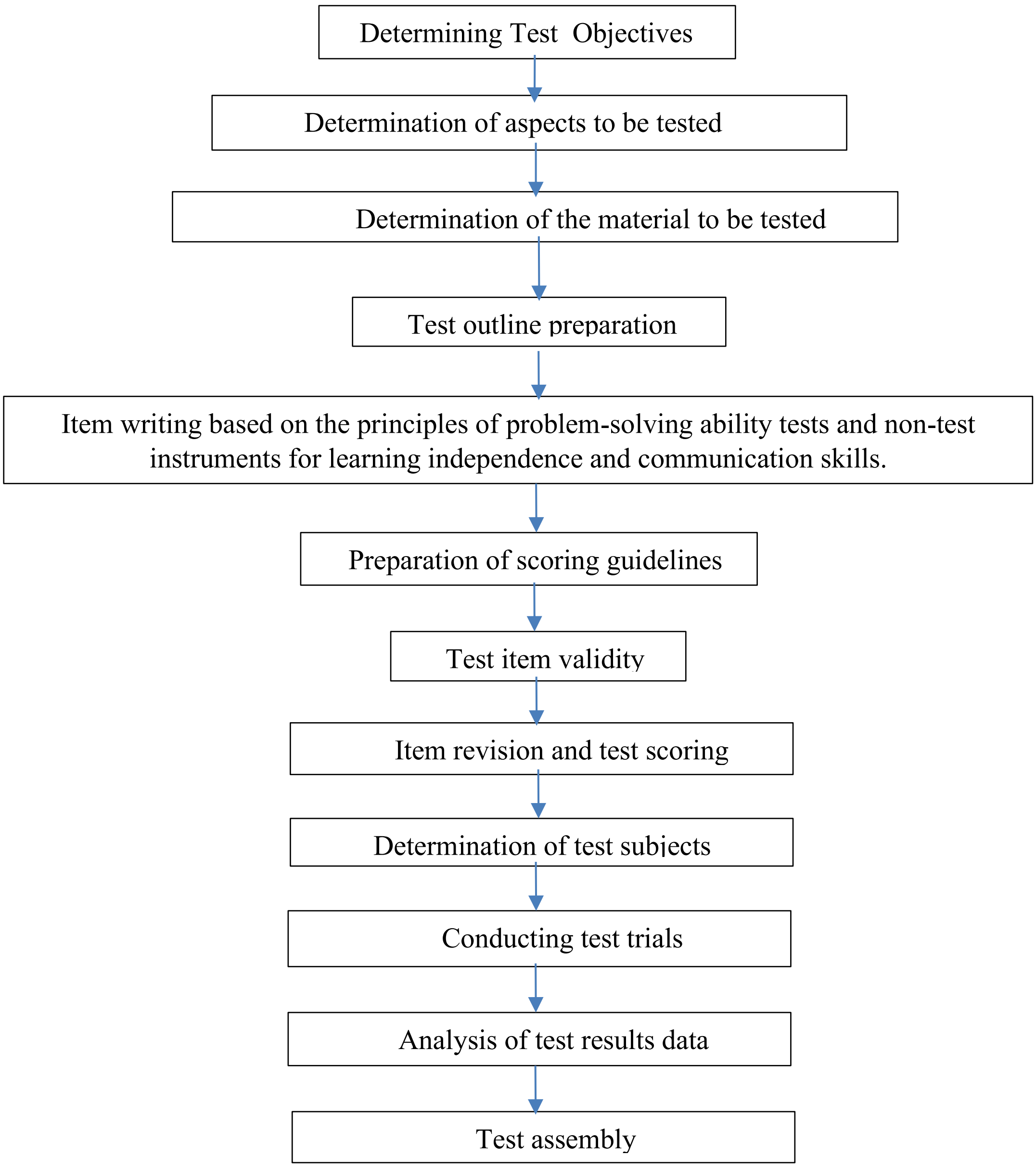

Figure 1. Flowchart developed

Research data analysis used the Rasch Model 1-PL for cognitive tests because the data were dichotomous, but for affective and psychomotor tests, the Graded Response Model was used. Data in this study were analyzed from several aspects, namely: item fit validity, reliability and SEM, difficulty index, and discrimination power.

In validity testing, especially construct validity, which is to ensure the suitability of test items and test participants with the model. Data obtained from the trial will be analyzed using statistical software. For cognitive instruments (dichotomous tests), the analysis will be performed with the Rasch Model using the Winsteps application. Meanwhile, for affective and psychomotor instruments (rating scales), the analysis will use the Graded Response Model (GRM) through the Parscale (PSL) application, using the Item Characteristic Curve (ICC). Item fit will be evaluated through the Mean Square Infit (MNSQ) value. If the INFIT MNSQ value for an item or test participant is outside the ideal range of 0.77 to 1.30 (i.e., greater than 1.30 or less than 0.77), the resulting ICC will appear flat. If the INFIT MNSQ value for an item or test participant exceeds 1.30 or is less than 0.77, the ICC will appear flat. This condition indicates that the distribution of the graph forms a platykurtic curve, no longer a leptokurtic curve(Keeves & Masters, 1999). Therefore, a test item or participant is considered to fit the model if its INFIT MNSQ value is within the range of 0.77 to 1.30 (Adam & Khoo, 1996).

According to classical test theory (CTT), reliability is defined as the accuracy or consistency of a test when performing measurements. This means that if a test is administered repeatedly to the same individual or group, the results should be relatively stable and not vary greatly (Crocker & Algina, 2008). One way to measure test reliability is to use Kuder-Richardson 20 (KR-20). The advantage of this method is that the test only needs to be administered once to respondents. After administering the single test, the reliability coefficient can be calculated using the KR-20 equation. In contrast, IRT uses test information functions to measure the quality of a test. These test information functions, in the context of IRT, play a role similar to that of reliability coefficients in CTT, although there are fundamental differences between the two (Istiyono, 2018). The test items in IRT consist of a set of items designed to form a single unit, from which the test information function can be obtained. Thus, in the context of IRT, test quality will be measured using the test information function, which plays a role similar to the reliability coefficient in CTT. For cognitive instruments (dichotomous tests), the analysis will be performed using the Rasch Model with the Winsteps application. Meanwhile, for affective and psychomotor instruments (rating scales), the analysis will use the Graded Response Model (GRM) through the Parscale (PSL) application.

Item difficulty analysis is used to evaluate the level of difficulty of questions, whether they are too easy or too difficult. An item is considered to be of good quality if it has a difficulty index value above -2.0 or below 2.0) (Baker, 2001; Hambleton R, 1985). Items that show a negative discrimination index (parameter a) should be removed from the test.n (Istiyono, 2018). This indicates an error in the item, where the probability of answering correctly decreases as the test-takers' ability increases. However, values greater than 2 are also uncommon. Therefore, the range of values generally considered acceptable is between 0 and 2 (Sumaryanta, 2021).

Data analysis techniques include how to interpret the data obtained, its relation to the problem, and the research objectives. There is no need to write statistical formulas for experimental research, but it is enough to state what tests were used and the decision-making criteria. For qualitative research, researchers must also describe what is done to ensure the validity and consistency of research results.

RESULTS & DISCUSSION

RESULTS

1. COGNITIVE TEST INSTRUMENT

The cognitive test instrument consists of 15 multiple-choice questions designed to measure problem-solving abilities. The results of the experts' synthesis are taken from (Bransford, 1993; Heller & Heller, 2010; Polya, 2004; Pólya, 2021) , then synthesized.

Subsequently, with expert judgment (practitioners and peer review), it was determined to be suitable for application.

Table 1. Distribution of problem-solving ability test items

Aspect	Sub-aspect	Number
Understanding the problem	Understanding	1
	Converting	2
	Comparing	3
Identifying the problem	Identifying	4,5
	Decomposing	6
	Analyzing	7
Selecting solutions	Solving	8
	Determining	9
Applying Solutions	Applying	10,11,12
	Deciding	13
Conducting Evaluation	Evaluating	14, 15

Table 2. INFIT MNSQ Values

Item	INFIT MNSQ	Criteria
1	0.99	Valid
2	0.80	Valid
3	1.00	Valid
4	1.00	Valid
5	0.88	Valid
6	1.23	Valid
7	0.86	Valid
8	0.91	Valid
9	0.94	Valid
10	0.96	Valid
11	0.85	Valid
12	1.27	Valid
13	1.19	Valid
14	0.97	Valid
15	1.17	Valid

According to the results of the trial conducted to measure the validity of the test items, INFIT MNSQ values were obtained for each item. The INFIT MNSQ values found ranged from 0.80 to 1.27. According to the criteria established by Adam & Khoo (1996). An item is considered suitable or valid for the model if its INFIT MNSQ value is within the range of 0.77 to 1.30. Thus, it can be concluded that all 15 items in the cognitive test instrument for measuring problem-solving ability have INFIT MNSQ values within the required range (0.77–1.30). This indicates that all items fit the model and are therefore considered valid in this measurement context.

Table 3. Classical Reliability Value

Reliability	Values	Category
Reliability alpha Cronbach	0,78	Reliable
Reliability item	0,91	Highly Reliable

Based on reliability testing using the Kuder-Richardson 20 (KR-20) method, a Cronbach's alpha value of 0.78 was obtained. Meanwhile, item reliability showed a value of 0.91. A Cronbach's alpha value of 0.78 is generally considered good, indicating that the test has adequate internal consistency. This means that the items in the test tend to measure the same construct consistently. The item reliability value of 0.91 is a very high number. This indicates that each item in the test individually has exceptional consistency in its measurement. Overall, these results indicate that the test used has strong reliability according to classical test theory. The internal consistency of the test and the stability of measurement for each item are excellent, making the scores obtained from this test reliable.

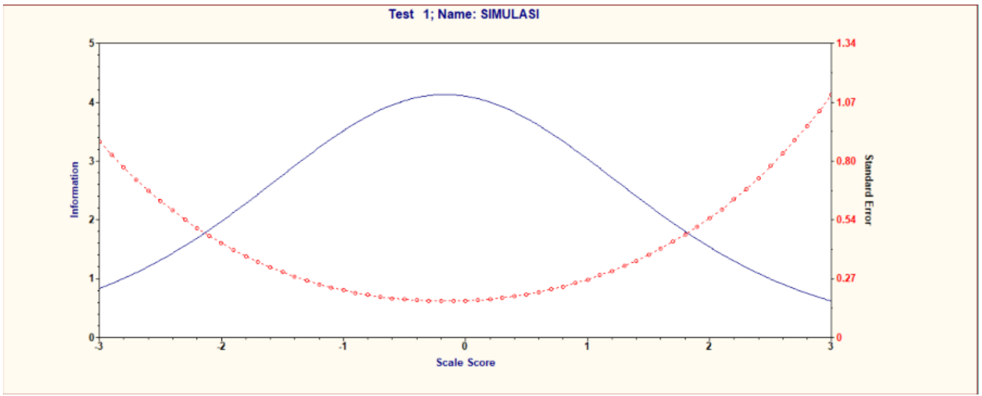

Figure 2. Information Function and SEM

Replacing the classical test theory, the test information function and standard error measurement (SEM) are used. Referring to the curve presented, this cognitive test is optimally designed to measure problem-solving abilities in students in the moderate category. This is indicated by the most effective ability range (θ), which is when $-2{,}15 \leq \theta \leq 1{,}85$.

Table 4. Item Difficulty Level

Item	MEASURE	Category
1	-1.12	Easy
2	0.92	Medium
3	-0.42	Medium
4	-0.21	Medium
5	0.53	Medium
6	0.80	Medium

Item	MEASURE	Category
7	-0.39	Medium
8	-0.52	Medium
9	0.15	Medium
10	0.32	Medium
11	-0.35	Medium
12	1.25	Hard
13	0.12	Medium
14	0.08	Medium
15	0.84	Medium

According to the analysis results, the difficulty level of the items in this cognitive test shows good quality. Referring to the criteria proposed by Baker (2001) and Hambleton R (1985). An item is of good quality if it has a difficulty index value above -2.0 or below 2.0. In this test, the range of difficulty levels found in the test items was from -1.12 to 1.25. Of the total test items, there was 1 item categorized as easy, 13 items categorized as moderate, and 1 item categorized as difficult. Therefore, it can be concluded that all items have difficulty levels consistent with the established standards and thus can be categorized as high-quality items for measuring cognitive problem-solving ability.

2. NON-TEST AFFECTIVE INSTRUMENTS

The effective non-test instrument consists of 15 questions with an intensity scale model, designed to measure learning independence. The synthesis of expert opinions was taken from (Arista & Kuswanto, 2018; Knowles, 1975; Leask & Pachler, 2014; Schunk & Zimmerman, 2013), then synthesized. Furthermore, based on expert judgment (practitioners and peers), it was deemed suitable for use.

Table 5. Distribution of items on the learning independence intensity scale

Aspect	Number
Initiative	1,2,3
Responsibility	4,5,6
Self-confidence	7,8,9
Use of Learning Resources	10,11,12
Evaluation	13,14,15

Table 6. INFIT MNSQ Value

Item	INFIT MNSQ	Criteria
1	0.87	Valid
2	1.16	Valid
3	1.14	Valid
4	0.86	Valid
5	1.03	Valid
6	0.93	Valid
7	1.18	Valid
8	1.01	Valid
9	1.16	Valid
10	0.97	Valid

Item	INFIT MNSQ	Criteria
11	0.88	Valid
12	0.87	Valid
13	1.00	Valid
14	0.92	Valid
15	1.08	Valid

Tests conducted have evaluated the validity of non-psychomotor test items in the form of an intensity scale designed to measure student learning independence. The analysis shows that the INFIT MNSQ values for each item range from 0.86 to 1.18. According to the criteria set by (Adam & Khoo, 1996), an item is considered suitable or valid for the model if its INFIT MNSQ value is between 0.77 and 1.30. Therefore, it can be concluded that all 15 items in this non-test instrument meet the required criteria. This indicates that all items fit the model used and are therefore valid in the context of measuring learning independence.

Table 7. Classical Reliability Value

Reliability	Values	Category
Reliability alpha Cronbach	0,90	Highly Reliable
Reliability item	0,78	Reliable

The reliability test conducted using the Kuder-Richardson 20 (KR-20) method produced a Cronbach's Alpha value of 0.90. This figure, 0.90, is generally categorized as highly reliable. This indicates that the test has a very good level of internal consistency, meaning that the items in it consistently measure the same construct or dimension. Additionally, an item reliability value of 0.78 was obtained. This value is considered high, indicating that each individual item in the test has good measurement consistency. Overall, these findings confirm that the test used has strong reliability according to the principles of Classical Test Theory. The high internal consistency of the test and the stability of measurement in each item make the scores produced by this test reliable.

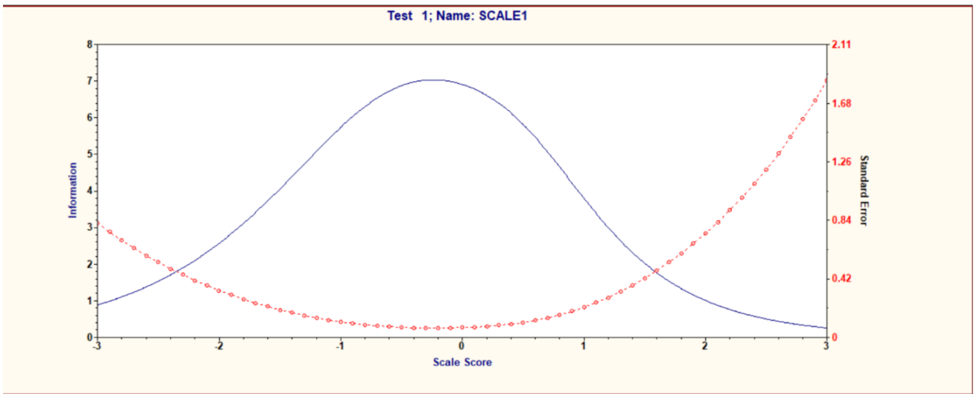

Figure 3. Information Function and SEM

In Item Response Theory (IRT), test information function and Standard Error of Measurement (SEM) serve as substitutes for the concept of reliability in classical test theory. Based on the curve presented, this intensity scale is optimally designed to measure learning independence in students of moderate ability. This is indicated by the most effective ability range (θ), which is when $-2{,}3 \leq \theta \leq 1{,}6$.

Table 8. Item Difficulty Level

Item	Measure	Criteria
1	-0.19	Medium
2	0.08	Medium
3	0.38	Medium
4	0.23	Medium
5	-0.09	Medium
6	-0.03	Medium
7	0.28	Medium
8	-0.13	Medium
9	-0.22	Medium
10	-0.01	Medium
11	0.03	Medium
12	0.20	Medium
13	0.02	Medium
14	-0.38	Medium
15	-0.17	Medium

Based on the analysis results, the level of difficulty of the items in this affective learning independence test shows good quality. In this test, the range of difficulty of the items found was from -0.38 to 1.28. Of the total items, all 15 items were categorized as moderate. Therefore, it can be concluded that all items have difficulty levels that align with the established standards. This indicates that these items are of good quality for measuring affective learning independence skills.

Table 9. Item Discrimination

Item	Slope
1	0,927
2	0.833
3	0.706
4	1.430
5	0.799
6	0.919
7	0.700
8	0.793
9	0.744
10	0.837
11	0.974
12	0.686
13	0.835
14	0.839
15	0.820

According to the analysis conducted, the item discrimination values obtained ranged from 0.686 to 1.430 on the non-test instrument measuring learning independence. This range indicates that all statement items have positive discrimination. In this context, the higher the discriminative power of an item, the more effective it is in distinguishing between individuals with high levels of learning independence and those with low levels of learning independence. This means that these items are highly effective in identifying differences in levels of learning independence among respondents. Thus, these results indicate that the non-test instrument has a strong ability to distinguish individuals based on their levels of learning independence.

3. NON-TEST PSYCHOMOTOR INSTRUMENTS

The psychomotor non-test instrument consists of 15 questions using a Likert scale, designed to measure communication skills. The results of the synthesis of expert opinions were taken from (DeVito, 2008; Oktaviani & Hidayat, 2015; Wood, 2008) , then synthesized. Furthermore, based on expert judgment (practitioners and peers), it was determined to be suitable for use.

Table 10. Distribution of non-test items on the Likert scale of communication skills

Aspect	Item
Expressing Opinions	1(+), 2(-), 3(+)
Clarity of Information	4(+), 5(-), 6(-)
Discussion Skills	7(+), 8(+), 9 (-)
Ethics	10(+), 11(+), 12 (-)
Listening	13(+), 14(-), 15 (-)

Table 11. INFIT MNSQ Value

Item	INFIT MNSQ	Criteria
1	0.84	Valid
2	0.82	Valid
3	1.06	Valid
4	0.90	Valid
5	0.93	Valid
6	1.17	Valid
7	0.85	Valid
8	0.83	Valid
9	1.10	Valid
10	1.12	Valid
11	0.86	Valid
12	1.19	Valid
13	1.14	Valid
14	0.93	Valid
15	1.17	Valid

The trial evaluated the validity of items on a non-psychomotor Likert scale designed to measure students' communication skills. The analysis showed that the INFIT MNSQ values for each item ranged from 0.82 to 1.19. Referring to the criteria set by Adam & Khoo (1996), An item is considered suitable or valid for the model if its INFIT MNSQ value is between 0.77 and 1.30. Thus, it can be concluded that all 15 items in this non-test instrument meet the required criteria. This indicates that all items are suitable for the model used and are therefore valid in the context of measuring communication skills.

Table 12. Classical Reliability Value

Reliability	Values	Category
Reliability alpha Cronbach	0.95	Highly Reliable
Reliability item	0.90	Highly Reliable

Reliability analysis using the Kuder-Richardson 20 (KR-20) method showed very positive results. The Cronbach's Alpha coefficient for this test reached 0.95, indicating an exceptionally high level of internal consistency. This means that all items in the test effectively and uniformly measure the same attribute. Furthermore, reliability at the item level is also very high, with a value of 0.90. This figure confirms that each item individually provides a highly consistent measurement. In short, based on the principles of Classical Test Theory, these results convincingly demonstrate that the test used has very strong reliability. The excellent internal consistency and measurement stability of each item ensure that the scores obtained from this test are highly reliable and trustworthy.

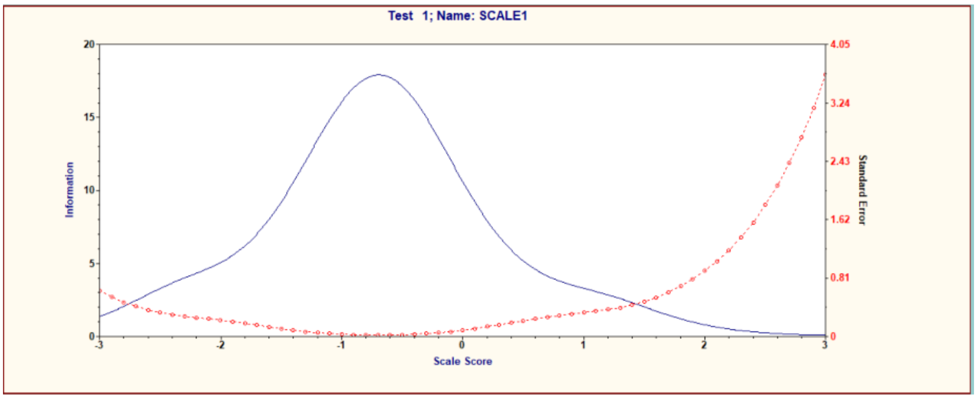

Figure 4. Information Function and SEM

The test information function and Standard Error of Measurement (SEM) in Item Response Theory (IRT) serve as quality metrics that replace the concept of reliability from Classical Test Theory. Based on the curve provided, it can be concluded that this Likert scale is most effective in measuring the communication skills of students with moderate abilities. The most effective ability range (θ) is when $-2{,}8 \leq \theta \leq 1{,}5$

Table 13. Item Difficulty Level

Item	Measure	Category
1	0.36	Medium
2	0.38	Medium
3	-0.12	Medium
4	-0.05	Medium
5	0.27	Medium
6	0.05	Medium

Item	Measure	Category
7	0.14	Medium
8	0.11	Medium
9	0.39	Medium
10	-0.97	Medium
11	-0.30	Medium
12	0.22	Medium
13	-0.93	Medium
14	0.19	Medium
15	0.38	Medium

Based on the analysis, the difficulty level of the items in this non-psychomotor communication skills test instrument shows good quality. In this instrument, the range of item difficulty found is from -0.97 to 0.38. Of the total 15 items, all are categorized as moderate. Therefore, it can be concluded that all items have difficulty levels that align with the established standards. This indicates that the items are of good quality for measuring psychomotor communication skills.

Table 14. Item Discrimination

Item	Slope
1	1.565
2	1.481
3	1.338
4	1.618
5	1.404
6	1.303
7	1.769
8	1.620
9	1.457
10	1.475
11	1.607
12	1.066
13	1.048
14	1.696
15	1.368

Analysis of non-test instruments designed to measure communication skills yielded significant findings regarding item discrimination. Item discrimination values ranged from 1.048 to 1.769. It is important to note that all statement items demonstrated positive discriminative power, a strong indication that each item contributes constructively to the measurement and no item functions inversely. Within the framework of Item Response Theory (IRT), discriminative power (parameter a or α) is a critical measure that describes how effectively an item can distinguish individuals based on their latent ability levels. Higher discriminative power values indicate that the item is more efficient in separating individuals with significantly different abilities. With the range of values found (1.048 to 1.769), the items in this instrument have excellent discriminative capability. Discrimination values above 1.0 are widely recognized as indicators of highly effective items in psychometrics, indicating that these items are accurate in identifying differences between individuals with high and low communication abilities

DISCUSSION

1. **Cognitive Instruments (Problem Solving)**
   Cognitive instruments are designed to measure students' problem-solving abilities in the context of bakpia production. The questions require students to apply their conceptual understanding to authentic scenarios often encountered in the bakpia-making process. For example, the aspects of "understanding the problem" and "identifying the problem" are tested through scenarios such as analyzing standard recipes and identifying potential errors in ingredient measurements, which are real problems in production. Questions about "applying solutions" and "evaluating" can be realized in practical challenges, such as calculating the proportion of ingredients to obtain the desired amount of dough or determining the optimal baking time to avoid production failures. The excellent quality of the questions, with an INFIT MNSQ value of 0.80–1.27, proves that these questions accurately measure students' problem-solving abilities in the context of bakpia.
2. **Non-Test Psychomotor Instruments (Communication)**
   Psychomotor instruments are designed to measure students' communication skills, which are a crucial aspect of teamwork when making bakpia. The aspects of "expressing opinions," "discussion skills," and "ethics" are tested through observation of interactions between students while collaborating, for example, when dividing tasks, discussing problems that arise, or providing feedback on the work of friends. Scenarios such as communicating ideas for flavor variations or explaining the reasons behind certain manufacturing methods are authentic means of measuring communication skills. The exceptionally high Cronbach's Alpha reliability value (0.95) and strong item discrimination (1.048–1.769) confirm that this instrument is consistently capable of measuring students' communication skills in a collaborative learning environment based on bakpia production activities.
3. **Non-Test Affective Instruments**
   Unlike cognitive and psychomotor instruments, affective instruments that measure learning independence are not directly integrated with bakpia material. The reason behind this is that learning independence is a broader construct and is part of an individual's character, not a skill that only appears when making bakpia. Learning independence encompasses initiative, responsibility, and self-evaluation, which are general and necessary in various aspects of life, both in school and outside of it. Therefore, the learning independence instrument is designed to measure how students act and think independently as a whole, without being tied to one specific activity. The integration of these three instruments—two context-based and one general—still provides a complete and holistic picture of student competence in accordance with the Pancasila learner profile. Nevertheless, the very high Cronbach's Alpha reliability value (0.90) and positive item discrimination values (0.686–1.430) indicate that this instrument is effective in distinguishing students' levels of learning independence.

CONCLUSION

This study successfully developed a set of integrated assessment instruments—for the cognitive, affective, and psychomotor domains—that have strong validity and reliability. The integration of local wisdom in the bakpia-making process proved to be an authentic, relevant, and meaningful context, making the measurement of 21st-century skills not only psychometrically valid but also contextually relevant to students.

The findings of this study have important implications for educational practice, namely that the instruments developed can be used by educators as a reliable tool for conducting holistic assessments of student competencies, including problem solving, independent learning, and communication, which are often not measured by traditional assessments. These results support a context- or culture-based learning approach that can increase student engagement and make abstract concepts easier to understand. This instrument can serve as a model for the development of similar assessments in other subjects.

Although this instrument shows very promising results, there is one limitation that needs to be noted. The contextualization of bakpia making, while a major strength, may pose a challenge if the instrument is applied in other regions. Therefore, further research is recommended to adapt this instrument to other relevant local contexts in various regions. This will ensure that the instrument has a broader scope and usefulness in comprehensively measuring student competencies.

REFERENCES

Adam, R. J., & Khoo, S. T. (1996). QUEST: The interactive test analysis system version 2.1. In *Australian Council for Educational Research*.

Alfika, Z. A., Mayasari, T., & Kurniadi, E. (2018). Pengembangan Instrumen Tes Kemampuan Memecahkan Masalah Fisika Berbasis Kearifan Lokal pada siswa SMP/MTs. *SEMINAR NASIONAL PENDIDIKAN FISIKA IV 2018*. http://e-journal.unipma.ac.id/index.php/snpf

Amir, I., Nursalam, N., & Mustafa, I. (2022). Tantangan Implementasi Nilai-Nilai Profil Pelajar Pancasila dalam Pembelajaran Bahasa Indonesia pada Kurikulum Merdeka Belajar. *GHANCARAN: Jurnal Pendidikan Bahasa Dan Sastra Indonesia*, 204–215. https://doi.org/10.19105/ghancaran.vi.7587

Angriani, A. D., Nursalam, N., Fuadah, N., & Baharuddin, B. (2018). PENGEMBANGAN INSTRUMEN TES UNTUK MENGUKUR KEMAMPUAN PEMECAHAN MASALAH MATEMATIKA SISWA. *AULADUNA: Jurnal Pendidikan Dasar Islam*, *5*(2). https://doi.org/10.24252/auladuna.v5i2a9.2018

Arista, F. S., & Kuswanto, H. (2018). Virtual physics laboratory application based on the android smartphone to improve learning independence and conceptual understanding. *International Journal of Instruction*, *11*(1). https://doi.org/10.12973/iji.2018.1111a

Asiati, S., & Hasanah, U. (2022). IMPLEMENTASI PROJEK PENGUATAN PROFIL PELAJAR PANCASILA DI SEKOLAH PENGGERAK. *Jurnal Lingkar Mutu Pendidikan*, *19*(2). https://doi.org/10.54124/jlmp.v19i2.78

Astuti, E. J., & Fitriani, R. (2021). Analisis Deskripsi Sikap Kemandirian Belajar Siswa Pada Mata Pelajaran Fisika Kelas X MIPA SMAN1 Bungo. *Journal Evaluation in Education (JEE)*, *2*(3). https://doi.org/10.37251/jee.v2i3.226

Audhiha, M., Vebrianto, R., Habibi, M., Febliza, A., & Afdal, Z. (2022). Pengembangan Instrumen Kemandirian Belajar untuk Siswa Sekolah Dasar. *Madrasah: Jurnal Pendidikan Dan Pembelajaran Dasar*, *14*(2). https://doi.org/10.18860/mad.v14i2.13187

Baker, F. B. (2001). The Basics of Item Response Theory. Second Edition. *Eric.Ed.Gov*.

Bransford, J. D. + S. B. S. (1993). *THE IDEAL PROBLEM SOLVER*. W. H. Freeman and Company.

Crocker, L., & Algina, James. (2008). Introduction to classical and modern test theory-Procedures for Estimating Reliability. In *Harcourt Brace Jovanovich College*.

DeVito, J. a. (2008). The Interpersonal Communication Book. *PsycCRITIQUES*, *32*.

Fitrianty, F., Yunita, A., & Juwita, R. (2022). Pengembangan Instrumen Tes Kemampuan Pemecahan Masalah Matematika di SMP Negeri 12 Padang. *Lattice Journal : Journal of Mathematics Education and Applied*, *2*(1). https://doi.org/10.30983/lattice.v2i1.5337

Griffin, P. + M. B. (2012). *Assessment and Teaching of 21st Century Skills* (E. Care, Ed.). Springer.

Hambleton R, S. H. (1985). Item Response Theory. Principles and applications. *Boston: Kluwer Nijhoff Publishing*.

Hasyim, M., Martawijaya, M. A., Helmi, H., Taqwin, M., Herman, N. M., & Wahid, S. (2022). *Pengembangan Instrumen Soal Berbasis Literasi Sains dan Kearifan Lokal pada Siswa Sekolah Menengah Atas* (SEMINAR NASIONAL HASIL PENELITIAN 2022, Trans.).

Heller, K., & Heller, P. (2010). Cooperative Problem Solving in Physics A User's Manual. *University of Minnesota*.

Istiyono, E. (2018). *PENGEMBANGAN INSTRUMEN PENILAIAN DAN ANALISIS HASIL BELAJAR FISIKA DENGAN TEORI TES KLASIK DAN MODERN*. UNY Press.

Istiyono, E., Mardapi, D., & Suparno, S. (2014). PENGEMBANGAN TES KEMAMPUAN BERPIKIR TINGKAT TINGGI FISIKA (PysTHOTS) PESERTA DIDIK SMA. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *18*(1). https://doi.org/10.21831/pep.v18i1.2120

Keeves, J. P., & Masters, G. N. (1999). Advances in Measurement in Educational Research and Assessment. In *Introduction*.

Knowles, M. (1975). *self-directed learning*. Follett Publishing Company.

Kurniawan, B. R., Reyza, M., Taqwa, A., & Fisika, P. (2018). Pengembangan Instrumen Tes Kemampuan Pemecahan Masalah Fisika pada Materi Listrik Dinamis. *Jurnal Pendidikan*, *3*(1).

Leask, M., & Pachler, N. (2014). *LEARNING TO TEACH USING ICT IN THE SECONDARY SCHOOL*. Routledge. http://www.bodleian.ox.ac.uk/education

Lesmana, C., & Nurussaniah, N. (2022). Integrasi Kearifan Lokal Kalimantan Barat dan ICT Berbasis Anroid dalam Media Pembelajaran Fisika. *Jurnal Basicedu*, *6*(2). https://doi.org/10.31004/basicedu.v6i2.2347

Maulida, F., & Sunarti, T. (2022). PENGEMBANGAN INSTRUMEN TES LITERASI SAINS BERBASIS KEARIFAN LOKAL DI KABUPATEN LAMONGAN. *ORBITA. Jurnal Hasil Kajian, Inovasi, Dan Aplikasi Pendidikan Fisika*, *8*(1).

Mery, M., Martono, M., Halidjah, S., & Hartoyo, A. (2022). Sinergi Peserta Didik dalam Proyek Penguatan Profil Pelajar Pancasila. *Jurnal Basicedu*, *6*(5). https://doi.org/10.31004/basicedu.v6i5.3617

Murti, W. W., & Sunarti, T. (2021). PENGEMBANGAN INSTRUMEN TES LITERASI SAINS BERBASIS KEARIFAN LOKAL DI TRENGGALEK. *ORBITA.Jurnal Hasil Kajian, Inovasi, Dan Aplikasi Pendidikan Fisika*, *7*(1).

Nisa, W. M., & Wilujeng, I. (2020). Teaching Science Books Based on Bakpia Local-Potential to Improve Students' Critical Thinking and Communication Skills. *Jurnal Penelitian Dan Pembelajaran IPA*, *6*(2). https://doi.org/10.30870/jppi.v6i2.8716

Nisa, W. M., & Wilujeng, I. (2021). Development of integrated science textbooks on local potential food Bakpia for junior high school. *Journal of Physics: Conference Series*, *1806*(1). https://doi.org/10.1088/1742-6596/1806/1/012126

Nugroho, A. S., & Mawardi, M. (2021). Pengembangan Instrumen Penilaian Sikap Tanggungjawab dalam Pembelajaran Tematik di Sekolah Dasar. *Jurnal Basicedu*, *5*(2). https://doi.org/10.31004/basicedu.v5i2.825

Nurhikmayati, I., & Sunendar, A. (2020). Pengembangan Project Based Learning Berbasis Kearifan Lokal Berorientasi pada Kemampuan Berpikir Kreatif dan Kemandirian Belajar. *Mosharafa: Jurnal Pendidikan Matematika*, *9*(1). https://doi.org/10.31980/mosharafa.v9i1.604

Nurkholifah, Mayasari, T., & Handhika, J. (2018). Pengembangan Instrumen Berpikir Kritis Berbasis Kearifan Lokal Pada Pembelajaran Fisika SMP. *SEMINAR NASIONAL PENDIDIKAN FISIKA IV 2018*.

Oktaviani, F., & Hidayat, T. (2015). PROFIL KETERAMPILAN BERKOMUNIKASI SISWA SMA MENGGUNAKAN METODE FENETIK DALAM PEMBELAJARAN KLASIFIKASI ARTHROPODA. *Jurnal Pengajaran Matematika Dan Ilmu Pengetahuan Alam*, *15*(1). https://doi.org/10.18269/jpmipa.v15i1.288

Partnership for 21st Century Skills. (2019). *Framework for 21st Century Learning*. Battelle For Kids.

Permata, V. W., Sahputra, R., Erlina*, E., Sartika, R. P., & Rasmawan, R. (2022). Hubungan Antara Kemandirian Belajar dengan Hasil Belajar Kimia Kelas X MIPA di SMA Negeri 6 Pontianak. *Jurnal IPA & Pembelajaran IPA*, *6*(4). https://doi.org/10.24815/jipi.v6i4.27418

Polya, G. (2004). *How to Solve It*. Princeton University Press.

Pólya, G. (2021). How to Solve It: ANew Aspect of Mathematical Method. In *A Century in Books: Princeton University Press 1905–2005*. https://doi.org/10.2307/2306109

Prabawanto, S. (2019). PENGEMBANGAN INSTRUMEN TES KEMAMPUAN PEMECAHAN MASALAH MATEMATIS SISWA. *Jurnal Pedagogik Pendidikan Dasar*, *6*(1). https://doi.org/10.17509/jppd.v6i1.21521

Realitawati, R. + I. F. D. + H. E. + K. D. (2024). PENERAPAN 4C SKILLS DALAM PEMBELAJARAN ABAD 21 DI SEKOLAH DASAR. *JURNAL MADRASAH IBTIDAIYAH*.

Rifa Hanifa Mardhiyah, Sekar Nurul Fajriyah Aldriani, Febyana Chitta, & Muhamad Rizal Zulfikar. (2021). Pentingnya Keterampilan Belajar di Abad 21 sebagai Tuntutan dalam Pengembangan Sumber Daya Manusia. *Lectura : Jurnal Pendidikan*, *12*(1). https://doi.org/10.31849/lectura.v12i1.5813

Rusnaini, R., Raharjo, R., Suryaningsih, A., & Noventari, W. (2021). Intensifikasi Profil Pelajar Pancasila dan Implikasinya Terhadap Ketahanan Pribadi Siswa. *Jurnal Ketahanan Nasional*, *27*(2). https://doi.org/10.22146/jkn.67613

Sae, F. S., Husin, V. E. R., & Mellu, R. N. K. (2021). Pengembangan Bahan Ajar Fisika Berbasis Kearifan Lokal Anyaman Nyiru untuk Meningkatkan Pemahaman Konsep Siswa. *Variabel*, *4*(1). https://doi.org/10.26737/var.v4i1.2321

Schunk, D. H., & Zimmerman, B. J. (2013). SELF REGULATION AND LEARING. In *HANDBOOK OF PSYCHOLOGY VOLUME 7: EDUCATIONAL PSYCHOLOGY*. John Willey & Sons.

Septiani, A. (2022). Implementasi kurikulum merdeka ditinjau dari pembelajaran matematika dan pelaksanaan P5 (studi di SMA Negeri 12 Kabupaten Tangerang). *AKSIOMA : Jurnal Matematika Dan Pendidikan Matematika*, *13*(3). https://doi.org/10.26877/aks.v13i3.14211

Sukmadana, I. W. A. N. W. (2024). Pembelajaran Berdiferensiasi Sebagai Upaya Menguatkan Profil Pelajar. *Prosiding Seminar Nasional Ilmu Pendidikan*, *1*. https://doi.org/10.62951/prosemnasipi.v1i1.17

Sumaryanta. (2021). *TEORI TES KLASIK DAN TEORI RESPON BUTIR*. CV. Confident (Anggota IKAPI Jabar).

Unik Hanifah Salsabila, Risqie Pramudya Endi, Roisul Ma'ruf, Sulaiman Saputra, & Inan Taat Diyanah. (2020). URGENSI TEKNOLOGI PENDIDIKAN DI ERA DISRUPSI. *Ed-Humanistics : Jurnal Ilmu Pendidikan*, *5*(2). https://doi.org/10.33752/ed-humanistics.v5i2.1074

Wood, J. T. (2008). Communication In Our Lives Fifth Edition. In *Wadsworth Cengage Learning*.

You, J. W., & Kang, M. (2014). The role of academic emotions in the relationship between perceived academic control and self-regulated learning in online learning. *Computers and Education*, *77*. https://doi.org/10.1016/j.compedu.2014.04.018