Research Article

Students' View on English Assessment in Indonesia's *Merdeka* Curriculum: A Kirkpatrick Evaluation Model

Nisa Khoiriyah*, Naf'an Tarihoran, Dini Fitriani

English Education, Postgraduate Program, Universitas Islam Negeri Sultan Maulana Hasanuddin Banten, 42118, Indonesia

ABSTRACT

Using Kirkpatrick's Four-Level Evaluation Model of reaction, learning, behavior, and results, this study investigates how students view English assessment in the Indonesian Merdeka Curriculum. This study involved 41 high school students who answered a valid and reliable Likert scale questionnaire, using a quantitative descriptive design. Students' perceptions were generally positive, according to descriptive statistical analysis conducted using SPSS 26. Most students acknowledged that the assessment was fair, relevant to classroom learning, and motivating. In addition, students reported behavioral changes such as increased classroom participation, but they experienced improvements in learning, especially in vocabulary and confidence. Although some students remained neutral, students said that the assessment helped their overall language development. The clarity of instructions, adequacy of time, and comfort during the assessment were also issues. These results indicate that, although the Merdeka Curriculum assessment method is in line with basic principles, some elements of practice need to be improved in order to optimize students' learning experiences and learning outcomes.

Keywords: English language assessment, Merdeka Curriculum, Kirkpatrick Model, student perception

Background

According to Brown (Brown, 2004), assessment is very important in language education because it serves as a measure of student achievement and as a tool to improve their learning process. The relatively new Indonesian Merdeka Curriculum includes assessment as an important component of learning. It is intended to be authentic, formative, and student-centered (The Ministry of Education, Culture, Research, 2024). This curriculum encourages teachers to use flexible, relevant, and meaningful assessment methods that allow students to actively participate in evaluating their progress.

However, in real life, the application of these assessment principles may not be consistent. This is because many teachers continue to use traditional methods due to limitations in training, resources, and clarity (Randall et al., 2022). According to Jannah & Rasyid (Jannah & Rasyid, 2023), teachers are also expected to take greater responsibility in designing innovative lessons and assessment tasks. In addition, Wuwur points out that there is a gap between what is expected in the curriculum and what happens in the classroom, especially in secondary schools (Wuwur, 2023).

Given these difficulties, it is crucial to understand how students view the assessment procedures they experience as direct recipients of the learning process. Understanding students' perspectives helps to identify differences between implementation and policy (Brookhart, 2023). Therefore, this study investigates students' perceptions of English assessment in a senior high school in Indonesia that uses the Merdeka Curriculum, using the four levels of Kirkpatrick's Evaluation Model: Reaction, Learning, Behavior, and Results.

Literature Review
English Language Assessment

According to Chapelle & Brindley (Chapelle & Brindley, 2019), assessment is a systematic process of collecting, analyzing, and interpreting information about students' language abilities in the four language skills: reading, writing, speaking, and listening. According to Brown (2004), effective evaluation must meet standards of validity, reliability, authenticity, and practicality. In contrast, Bachman and Palmer (1996) note that the key qualities of language tests are interactivity, impact, and fairness. The purpose of these principles is to ensure that language tests not only measure students' memorization skills, but also actively encourage them to learn, provide feedback, and identify learning gaps. As a result, when well-designed, language assessment helps students learn independently and make instructional decisions.

Assessment Practices in the Merdeka Curriculum

Merdeka Curriculum Assessment Methods The Merdeka Curriculum was created in response to learning loss after the pandemic and aims to improve student literacy and skills through a flexible, student-centered approach to learning (Intiana et al., 2023). English language education in this curriculum is essential for teaching students academic engagement, cultural understanding, and global communication skills (Masita, 2024). English language teaching is based on six integrated macro skills: reading, watching, speaking, writing, listening, and presenting (The Ministry of Education and Culture, 2022b). Each skill is taught to enhance critical and creative thinking, technological literacy, and self-expression (The Ministry of Education and Culture, 2022a).

The Merdeka Curriculum incorporates assessment as an important component of learning to support these goals. According to Samsudi (Samsudi, 2023), assessment is intended to be formative, authentic, and flexible. It emphasizes a growth mindset, integration between the affective and cognitive domains, and responsiveness to individual learning needs. Learning objectives, teaching strategies, and task types must be aligned for effective assessment design (Anderson et al., 2001). By prioritizing fairness, proportionality, validity, and reliability in the free curriculum, teachers are given the freedom to choose assessment methods and timing (Anggreana et al., 2024). This curriculum encourages educators to assess the learning process and learning outcomes. The results of this assessment can be used to make teaching decisions, provide feedback, identify students' strengths and interests, and support character development (Samsudi, 2023).

his chair aims to support the implementation of authentic assessment. According to Mueller (Mueller, 2005), this type of assessment involves students in real and meaningful tasks that demonstrate the use of language outside the classroom, such as conducting interviews, giving multimedia presentations, or working on collaborative projects. However, the implementation of this type of assessment

is still limited in practice. Many educators face difficulties in implementing authentic assessment due to perceived complexity, time constraints, lack of training, and lack of institutional support. This occurs instead of understanding the concept (Lismawati et al., 2023).

Kirkpatrick's Evaluation Model

Kirkpatrick's Four Levels of Evaluation (Kirkpatrick & Kirkpatrick, 2007) provides a comprehensive framework for evaluating the effectiveness of educational practices, including assessment. This model has four levels: Reaction, Learning, Behavior, and Results. At the first level, the evaluation focuses on participants' immediate reactions to the assessment experience. At the second level, learning measures how much participants have gained the desired knowledge, skills, and attitudes. At the third level, behavior examines whether participants' behavior has changed outside the classroom. At the final level, evaluation focuses on how participants have gained knowledge.

Methods

This study uses a quantitative descriptive design to examine how students view English assessment. This method uses Kirkpatrick's Four Levels of Evaluation Model to conduct this investigation.

This study was conducted at Al-Azhar 6 Islamic High School in Cilegon. Forty-one students from grades X and XI were involved. Convenience sampling was used to select participants based on their availability and interest in participating. The findings may not be applicable to a broader context due to the small sample size (N = 41).

This study used a Likert scale questionnaire based on the four levels of the Kirkpatrick Model: Reaction, Learning, Behavior, and Results. The values were 1 for "Strongly Disagree," 2 for "Disagree," 3 for "Neutral," 4 for "Agree," and 5 for "Strongly Agree." There were 24 items in the questionnaire that showed indicators aligned with the model. The results of the instrument's validity and reliability tests show that it is valid and reliable for this study, with Cronbach's Alpha of 0.866, 0.832, 0.798, and 0.689, respectively. Furthermore, construct validity was confirmed by conducting item-total correlation analysis using SPSS 26, with acceptable correlation coefficients for each item.

Data were collected during the even semester of the 2024/2025 academic year. With the approval of the school and teachers, participants completed the survey during class hours. To interpret students' responses to each item in the four levels of Kirkpatrick's framework, descriptive statistical analysis was performed using SPSS version 26. The focus of the analysis was on frequency and percentage.

Result and Discussion
Level 1 Reaction

The first level of evaluation focuses on how students respond to the assessment. Although there were many neutral responses, students generally responded positively to the assessment. Only 46.3% of students agreed or strongly agreed that the instructions were clear, while 41.5% remained neutral. With regard to the assessment tasks, the same trend was observed: 46.4% agreed, while 48.8% were neutral. This neutrality may indicate indifference, that some students were not very engaged, or that the assessment experience was not important enough to form a strong opinion. Furthermore, neutrality may stem from ambiguity in the assessment design itself, particularly in learning objectives and instructions. This ambiguity may not have been clearly communicated, causing students to be unable to clearly agree or disagree. This neutrality indicates that learning objectives in assessments need to be better communicated and aligned.

Students consider assessment to be generally fair (58.5% agree), helpful in identifying their strengths and weaknesses (58.5% agree), and motivating (70.8%). Returning to neutrality, responses to the comfort level of assessment and adequacy of time varied. The item on time adequacy is the most concerning, with 17.1% of respondents expressing dissatisfaction. Overall, 51.2% are satisfied with the assessment experience, while 48.8% say they are neither satisfied nor dissatisfied. These results indicate that, although the assessment experience is moderately positive, there is still room for improvement in terms of clarity, comfort, and time for students.

Table 1. Students' Responses to English Language Assessment – Level 1: Reaction

No.	Topic	SD (%)	D (%)	N (%)	A (%)	SA (%)
1	Instructions were clear	0.0	4.9	41.5	26.8	19.5
2	Tasks were relevant	0.0	4.9	48.8	17.1	29.3
3	Fair opportunity to show skills	0.0	7.3	34.1	39.0	19.5
4	Helped identify strengths/weaknesses	0.0	9.8	31.7	31.7	26.8
5	Motivated to do best	0.0	2.4	26.8	41.5	29.3
6	Felt comfortable	9.8	7.3	39.0	24.4	19.5
7	Sufficient time	9.8	7.3	39.0	24.4	19.5
8	Overall satisfaction	0.0	0.0	48.8	34.1	17.1

Level 2 Learning

Secondary level learning evaluates whether students feel an improvement in their knowledge or skills. The results show that the outcomes are generally positive. For example, 48.8 percent of respondents reported a better understanding of grammar, and 56.1 percent reported that their vocabulary had improved. In addition, most people (68.3%) were aware of areas where they could improve themselves, and 68.3% also said they were more confident in speaking English.

They were slightly less interested in real-life context application, learning feedback, and verbalization of understanding. Neutrality remained high across these indicators, suggesting that students were either uncertain about the role of assessment in supporting these areas or experienced varied degrees of impact. Nevertheless, most students (over 50%) acknowledged the learning benefits from the assessment, suggesting that it fulfilled its formative function reasonably well.

Table 2. Students' Responses to English Language Assessment – Level 2: Learning

No.	Topic	SD (%)	D (%)	N (%)	A (%)	SA (%)
9	Understanding grammar	0.0	4.9	46.3	26.8	22.0
10	Improved vocabulary	0.0	0.0	43.9	41.5	14.6
11	Identify areas for improvement	0.0	4.9	26.8	41.5	26.8
12	More confident in English	0.0	2.4	29.3	41.5	26.8
13	Apply English in real life	2.4	7.3	39.0	39.0	12.2
14	Feedback helped learning	2.4	4.9	46.3	36.6	9.8
15	Can explain concepts to others	2.4	2.4	46.3	41.5	9.8

Level 3 Behavior

The third level of evaluation examined whether students' learning experiences were transferred into observable behavior, such as increased participation and language use beyond assessment settings. Findings at this level were more varied. Items such as increased engagement in English, both in and outside the classroom, received relatively low agreement rates. Only 21.9% of students agreed that they used English more actively during class discussions, while 24.4% reported using English in informal contexts outside the classroom.

Nevertheless, some indicators showed more promising responses. A notable 63.4% of students stated that they were more likely to use English learning resources independently, and 46.3% reported that they made more effort to use complex grammar and vocabulary in their writing. These results suggest that while the assessment did not strongly influence general language behavior, it did encourage some students to seek out additional resources and apply more advanced language use in specific contexts.

Table 3. Students' Responses to English Language Assessment – Level 3: Behavior

No.	Topic	SD (%)	D (%)	N (%)	A (%)	SA (%)
16	More active in class discussions	7.3	31.7	39.0	19.5	2.4
17	Use English outside class	12.2	34.1	29.3	19.5	4.9
18	Use more English resources	0.0	9.8	26.8	36.6	26.8
19	Use complex grammar/vocab	2.4	17.1	34.1	31.7	14.6
20	Confident in participation	2.4	12.2	53.7	24.4	7.3
21	Seek speaking practice	2.4	9.8	31.7	39.0	17.1

Therefore, there is a clear contrast between the results at Level 2 Learning and Level 3 Behavior. While the assessment appeared to support students' internal learning outcomes—such as improved vocabulary (56.1%), increased confidence (68.3%), and better self-awareness of their strengths and weaknesses (68.3%)—these perceived gains did not fully in line with behavioral change. Only 21.9% of students reported being more active in class discussions, and 24.4% used English more frequently outside the classroom.

This distinction suggests that the Merdeka English assessment seems to encourage students' private learning behaviors—for instance, seeking additional English learning resources (63.4%) or attempting to use complex grammar and vocabulary (46.3%)—but it falls short in developing socially interactive behaviors, such as speaking practice, collaboration, or public participation.

Such imbalance has important implications for the Merdeka Curriculum, whose core principles emphasize active learning, collaboration, and communication.

Level 4 Result

The fourth level focused on students' perceived outcomes following the assessment. The findings indicate a generally positive perception of the language development. Over half of the students (53.7%) agreed that their communication skills improved as a result of the assessment. Similarly, 48.8% acknowledged that the assessment helped them express their ideas more clearly, and 48.7% felt their overall English proficiency had improved.

Despite these positive indicators, most students responded neutrally to these items, suggesting that the perceived impact of the assessment on language mastery was moderate. These results project the partial effectiveness of the assessment in producing meaningful impact. Therefore, they suggest that further instructional support and post-assessment reinforcement may be needed to maximize learning outcomes.

Table 4. Students' Responses to English Language Assessment – Level 4: Result

No.	Topic	SD (%)	D (%)	N (%)	A (%)	SA (%)
22	Improved communication skills	2.4	2.4	41.5	41.5	12.2
23	Can express ideas creatively	0.0	9.8	41.5	39.0	9.8
24	Improved overall proficiency	2.4	4.9	43.9	34.1	14.6

Conclusion

Using Kirkpatrick's Four-Level Evaluation Model, this study investigated how students view English assessment in a free curriculum. Findings show that students generally respond positively to assessment, particularly in terms of its effect on their motivation, their perception of fairness, and its impact on their English comprehension. Most students agreed that assessment encouraged them to perform well and improve their grammar, vocabulary, and self-awareness at the reaction and learning levels.

However, the impact was less apparent in behavior and results. Many students did not report significant changes in their classroom participation or their use of English outside of the academic environment. Although more than

half of the students said they had improved their communication and communication skills, many neutral responses indicated that not all students fully understood this information.

This imbalance has significant consequences for the Merdeka Curriculum, which focuses on active learning, teamwork, and communication. The results show that although the assessment meets the formative learning objectives, it may not meet the broader social objectives of the curriculum, which are to make students more confident and communicative in English. Performance-focused and interaction-oriented tasks that recognize verbal participation and correct language use in communication may need to be incorporated into future assessment designs.

This study emphasizes the importance of creating valid and relevant assessments and ensuring that feedback, instructional support, and follow-up activities are integrated to improve learning outcomes.

Recommendations

Based on the results of the four levels of evaluation, several specific actions are proposed to improve the implementation of English assessment in the Merdeka Curriculum.

1. Enhance Clarity and Time Management

Teachers should use clear rubrics and time allocations that are consistent with the assessment objectives. Organized scheduling and transparent standards can help students view assessment tasks with more confidence and meaning as they reduce uncertainty and negative perceptions.

2. Strengthen Behavioral Changes through Communicative Tasks

Assessment has helped improve individual learning, but it is not enough to encourage communication. Post-assessment activities such as collaborative projects, peer-led discussions, and short presentations should be included. With the help of these activities, assessment can be transformed from a mere evaluation process into an opportunity to use language correctly. This can enhance the social goals and participation of the Merdeka Curriculum.

3. Deepen Contextual Understanding through Qualitative Inquiry

Future research should use qualitative approaches, such as focus group discussions or semi-structured interviews, to identify the reasons why students tend to be neutral and transform limited learning into active participation in class. By conducting such research, richer contextual insights will be provided and quantitative evidence will be supplemented. This will enable more informative and responsive assessment designs to be implemented in the future.

References

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., Raths, J., & Wittrock, M. C. (2001). A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives (L. W. Anderson & D. R. Krathwohl (eds.)). Longman.

Anggreana, Y., Ginanto, D., Felicia, N., Andiarti, A., Herutami, I., Alhapip, L., Iswoyo, S., Hartini, Y. M., & Rizal Listyo. (2024). Panduan Pembelajaran dan Asesmen Pendidikan Anak Usia Dini, Pendidikan Dasar, dan Menengah. Ministry of Education, Culture, Research, and Technology. https://kurikulum.kemdikbud.go.id/tentang

Bachman, L. F., & Palmer, A. S. (1996). Language Testing in Practice: Designing and Developing Useful Tests. Oxford University Press.

Brookhart, S. M. (2023). Assessment Literacy in A Better Assessment Future. Chinese Journal of Applied Linguistics. https://doi.org/10.1515/CJAL-2023-0202

Brown, H. D. (2004). Language Assessment: Principles and Classroom Practices. Pearson Education.

Chapelle, C. A., & Brindley, G. (2019). Assessment. In N. Schmitt (Ed.), An Introduction to Applied Linguistics. Hodder Education. https://doi.org/10.4324/9780429424465-17

Intiana, S. R. H., Prihartini, A. A., Handayani, F., Mar'i, M., & Faridi, K. (2023). Independent Curriculum and the Indonesian Language Education throughout the Era of Society 5.0: A Literature Review. AL-ISHLAH: Jurnal Pendidikan, 15(1), 911–921. https://doi.org/10.35445/alishlah.v15i1.3140

Kirkpatrick, D. L., & Kirkpatrick, J. D. (2007). Implementing The Four Levels: A Practical Guide for Effective Evaluation of Training Programs (First). Berrett-Koehler Publishers, Inc.

Lismawati, Wasehudin, Naf'an Tarihoran, Fitri Hilmiyati, & Fadilla Oktaviyana. (2023). Implementation of Authentic Assessment of the Implementation of Merdeka Curriculum in PAI Subjects at. TADRIB: Jurnal Pendidikan Islam, 9(1), 174–185. https://doi.org/10.19109/tadrib.v9i1

Masita, E. (2024). Buku Referensi-Pengajaran Bahasa Inggris dalam Kurikulum Merdeka. PT Media Penerbit Indonesia.

Mueller, J. (2005). The Authentic Assessment Toolbox: Enhancing Student Learning through Online Faculty Development. Journal of Online Learning and Teaching, 1(1).

Randall, R., Sukoco, G. A., Heyward, M., Purba, R., Arsendy, S., Zamjani, I., & Hafiszha, A. (2022). THE LEARNING GAP SERIES-TWO Reforming Indonesia's Curriculum: How Kurikulum Merdeka Aims to Address Learning Loss and Learning Outcomes in Literacy and Numeracy. INOVASI. https://www.inovasi.or.id

Samsudi. (2023, November 1). Konsep Asesmen dalam Kurikulum Merdeka. (Samsudi, Performer) Online National Training. https://www.youtube.com/watch?v=ejNxQ3P_In0

The Ministry of Education, Culture, Research, and T. (2024). Regulation of the Minister of Education, Culture, Research, and Technology of the Republic of Indonesia Number 12 of 2024. jdih.kemdikbud.go.id

The Ministry of Education and Culture. (2022a). Capaian Pembelajaran Mata Pelajaran Bahasa Inggris Fase A-Fase F. Ministry of Education, Culture, Research, and Techology Indonesia.

The Ministry of Education and Culture. (2022b). Tanya Jawab Kurikulum Merdeka. The Ministry of Education and Culture. https://repositori.kemdikbud.go.id/25344/

Wuwur, E. S. P. O. (2023). Problematika Implementasi Kurikulum Merdeka di Sekolah Dasar. Jurnal Ilmu Pendidikan, 3(1), 1–9. https://ejurnal.politeknikpratama.ac.id/index.php