Development of character assessment instrument on English learning for middle school students

Andi Abdurrahman Manggaberani [a,1,*]; Nur Hidayanto Pancoro Setyo Putro [b,2]

[a] Educational Research and Evaluation, Post-Graduate Program, Universitas Negeri Yogyakarta, Jl. Colombo No.1, Karang Malang, Caturtunggal, Sleman, Special Region of Yogyakarta 55281, Indonesia

[b] English Language Education Study Program, Faculty of Languages, Arts and Culture, Universitas Negeri Yogyakarta, Jl. Colombo No.1, Karang Malang, Caturtunggal, Sleman, Special Region of Yogyakarta 55281, Indonesia

[1] andiabdurrahman.2021@student.uny.ac.id*; [2] nur_hidayanto@uny.ac.id

* Corresponding author

Abstract: There is a present-day necessity to construct character education assessment instruments in English courses. Employing the ADDIE development model, this research undergoes stages encompassing analysis, design, development, implementation, and evaluation to craft a meticulously designed assessment instrument. The resulting instrument is a self-assessment for English subjects, subjected to scrutiny through both Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA). The instrument attests to content validity, indicated by an Aiken index of 0.955, and construct validity using EFA & CFA. Outcomes showcasing loading factor values exceeding 0.4 for each item and the model for measuring student character aligns seamlessly with the hypothesized model, substantiated by statistical results of p-values = 0.056; CFI = 0.979; TLI = 0.972; RMSEA = 0.031; SRMR = 0.055. Furthermore, the item reliability in each dimension/factor surpasses 0.7. The study's revelations uncover that 46% of students predominantly concentrate on character development, while 35% prioritize character emphasis. Cultivated character for 14%, whereas a mere 5% of students necessitate guidance. This instrument not only guides educators and researchers in developing character assessment tools but also meets valid and reliable criteria, ensuring credibility and usability across various education levels.

Keywords: assessment; character; English learning; instrument development

1. Introduction

The pivotal role of student attitudes in shaping the overarching objectives of learning, particularly their dispositions towards the target language, holds significance in educational contexts (Harmanto et al., 2023). Character education is the process of learning about the internal resources that each person can use to achieve and live a moral, rewarding, and productive life (Tirrell, 2022). The cultivation of personal attributes holds significant importance, particularly in early childhood, as the establishment of a robust groundwork is imperative for fostering constructive and supportive character development.

It is necessary to know that strong character consists of elements such as moral knowledge, moral feelings, and moral actions (Darnell et al., 2019). Instilling character education is important for today's young generation because character education aims to prepare them to face global competition by instilling strong character values. According to Afifuddin and Burga (2022), character education shares essential similarities with moral education. The overarching aim is to delineate children's personalities, fostering their development into responsible individuals who contribute positively to society and the state. This aligns with Jabbarova's (2022) perspective emphasizing responsible and positive interactions with others. Furthermore, Dariyanto and Awiria (2023), highlight that the primary objective of designing learning programs is to shape positive character traits, with a focus on aligning these programs with the various stages of child development.

In investigating the diverse perspectives on character building by scholars, character education is an educational initiative that collaborates with various societal entities to shape students' character systematically and explicitly. It is grounded and aims to

influence the behavior of the younger generation, ultimately seeking to restore their moral consciousness (García-Moriyón et al., 2020; Arif et al., 2023). To achieve the ideal goals of character education, the active involvement of all relevant entities is crucial. Mere discussions and conceptualization are insufficient without a comprehensive action plan that includes all components. The objective of character education is to empower students to enhance and apply their knowledge independently and manifesting these principles in their daily behavior.

However, based on the results of a survey conducted by the Indonesian Child Protection Commission (KPAI R.N, 2021), between 2016 and 2020, there were 2626 cases of children engaging in criminal activities. This prevalent social phenomenon underscores the importance of enhancing character education for students. The high incidence of violence, involving both victims and perpetrators among children, emphasizes the urgent need for improved character development in education. The Special Region of Yogyakarta shows a significant level of criminality, especially among children and adolescents (Wijayanti & Pratomo, 2019). This situation raises concerns because Yogyakarta is known as an educational center that attracts many students and students from various parts of Indonesia who choose to continue their education (Yu & Setiyaningrum, 2018). The study results from Marela et al. (2017), uncovered a notable connection between bullying victimization and a heightened prevalence of depression among high school in Yogyakarta. This prevalence underscores serious concerns about the safety and well-being of children and emphasizing the urgent need for intervention and support.

The role of education is acknowledged as a proactive measure in fostering a forthcoming generation that contributes to the enhancement of a superior society (Nisa et al., 2021). Education is anticipated to enhance the multifaceted qualities of the nation's populace, thereby mitigating, and diminishing the origins of diverse character issues. Within educational institutions, a transformative alteration in the instructional approach is imperative, encompassing a focus on character education within the learning framework (Susanto et al., 2022). The significance of character education resides in its capacity to cultivate enhanced individuals within society, thereby establishing circumstances conducive to increased societal advantages and greater prospects for individual success. The incorporation of character education into the English language learning process plays a pivotal role in cultivating students proficient in English communication, both verbally and in writing, while simultaneously imbuing and shaping their character in alignment with the goals outlined in Indonesia's national education agenda

In the contemporary epoch characterized by swift global interconnectedness, English can be identified as the predominant global lingua franca, rendering it the most extensively utilized language across diverse geographical locales worldwide (Rao, 2019; Tahmasbi et al., 2019; Liando & Tatipang, 2022). As a result of this impact, education systems in various countries require English classes from the secondary or even primary level (Koenig et al., 2020). The role of an educator assumes a significant role by making an essential contribution. This is because teachers guide students in achieving holistic competence in English.

Sumague (2020) emphasizes that a proficient language educator should possess key attributes such as employing experimental approaches and adapting teaching styles for effective knowledge transmission. The significance of English as a foreign language in Indonesia's curriculum presents a challenge for teachers to develop impactful and innovative lessons within the evolving national curriculum. Sofiana et al. (2019) underscore the responsibility of English educators to conduct comprehensive assessments covering affective, cognitive, and psychomotor dimensions. The goal is to nurture students' creative and critical abilities, as discussed by Malaikosa and Taopan (2020).

Researchers have created various assessment instruments to evaluate students' character, with studies such as Hidayati et al. (2018), Alivi and Widihastuti (2022), and Sadri et al. (2020) developing tools for student responsibility, affective self-assessment in chemistry, and character-based assessment in Indonesian language subjects, respectively. Notably, the development of English language mastery in schools is a very crucial policy in

this era of globalization because English is an international language that opens various opportunities, including in the education sector. Through the English learning process, students are expected to have the ability to master communication skills proficiently. Because in the context of communication, language functions as an effective tool to interact with other individuals (Akhter et al., 2020).

The role of English if used incorrectly by students, especially in verbal bullying or cyberbullying has a very dangerous impact (Ibrahim et al., 2024). This is because if students from the beginning of their educational journey use English incorrectly, then the possible outcomes in the long run can threaten the quality of their education and social interactions. In the context of English language learning, students are not only taught about the linguistic structure of the language but also given guidance on how their character can grow and develop through interactions in the classroom (Heritage et al., 2020). Students' characters can be shaped or nurtured through English language teaching in the classroom. The assessment of students' character at the junior high school level has great significance considering that the learners are experiencing a phase change from elementary education to junior high school and have entered the early adolescent phase.

Therefore, it is important to manage students' behavior early on so that schools and parents can provide reports on their development in various aspects such as academic achievement, social interaction, emotional well-being, and their actions. One way to evaluate students' abilities to handle English subjects is through the use of appropriate evaluation tools. Mardapi (2017) emphasizes that instrument quality hinges on validity and reliability. The character assessment instrument development process integrates relevant theories, adhering to evaluation principles like validity, reliability, fairness, and accuracy. Balancing theoretical foundations with local contexts and stakeholder input is crucial for crafting effective assessment tools.

The validity of data collection tools or instruments is the most complicated and most important thing to consider when conducting or criticizing a study (Vogt, 2018). Instrument validity refers to the extent to which the instrument precisely measures the dimensions it is supposed to measure or collects accurate data (Kleemann et al., 2023; Zhang & Garcia, 2023). There are three general or main types of validity traditionally consisting of content validity, construct validity, and criterion validity (Heale & Twycross, 2015; Yusuf, 2015; Azwar, 2021).

This study underscores the significance of content validity in the creation of research instruments. Content validity entails the adjustment and alignment of questions or instrument items with the targeted constructs or variables for measurement. It is a validity type confirmed through a rational analysis conducted by a panel of experts, as outlined by Istiyono (2020). Content validity ensures that the measurement process includes elements that are sufficient and representative, fully capturing the concept being assessed. Furthermore, construct validity is very suitable for measuring something latent or the framework of a concept, this is because construct validity involves empirical evidence that supports that the instrument validly reflects or describes the intended construct. According to Retnawati (2016), proving construct validity can be done in two ways, the two ways are Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA).

EFA is a fundamental tool in the development and validation of psychological theory and measurement (Watkins, 2018). The primary goal of factor analysis is to condense information, making it easier to describe and comprehend relationships and patterns (Yong & Pearce, 2013; Hair et al., 2019; Panahi et al., 2023; Steenkamp & Olivares, 2023). While CFA is a statistical method utilized to evaluate the validity of a previously proposed or established factor structure or measurement model (Brown, 2015; Brauer et al., 2023; Panahi et al., 2023; Yusoff et al., 2023). In instrument development, it is necessary to assess the validity and reliability of the measuring instrument to be used. If a measuring instrument is not valid or reliable, then the test may not accurately reflect the desired behavior or quality. Data collection tools are instruments used in research to gather information systematically, and the concepts of validity and reliability are closely related in an instrument in research, although they express different properties (Amalia et al., 2023).

The research aims are to delve into the profound significance of character development in tackling the intricate challenges posed by the contemporary world. The study endeavors to contribute to the enrichment of our comprehension of individual character, aspiring to craft a valuable instrument that can significantly augment character education initiatives in Indonesia. The research has been meticulously designed to adhere to rigorous criteria of validity, reliability, and effectiveness. The guiding research questions encompass the construction of a character education assessment instrument tailored for English subjects among junior high school students. Additionally, the study aims to assess the quality of the developed instrument and evaluate the measurement results obtained through its application in character education assessment for junior high school students studying English. Through this comprehensive investigation, the research seeks to provide insights and tools that can contribute meaningfully to the advancement of character education practices in the Indonesian educational landscape.

2. Materials and Methods

This study constitutes developmental research, commonly recognized as Research and Development (R&D), employing the ADDIE model proposed by Branch (2009). The ADDIE model offers a systematic instructional design framework, integrating an iterative process with crucial developmental steps (Aydin et al., 2023). The outcome of this investigation is a product manifested as a character assessment tool, specifically a Self-Assessment in English subjects tailored for junior high school students, developed utilizing a modified ADDIE development model. The research process for developing character assessment instruments in this study uses the ADDIE model proposed by Branch (2009), and has been combined and modified based on references put forward from Jurianto (2017); Nisa et al. (2021); Yoong (2022); Feng & Sangsawang (2023); Idris & Rus (2023); Nurmiati et al. (2023); Pulukadang et al. (2023); Shakeel et al. (2023). The flow of research stages in the development of character instruments can be explained briefly in Figure 1.
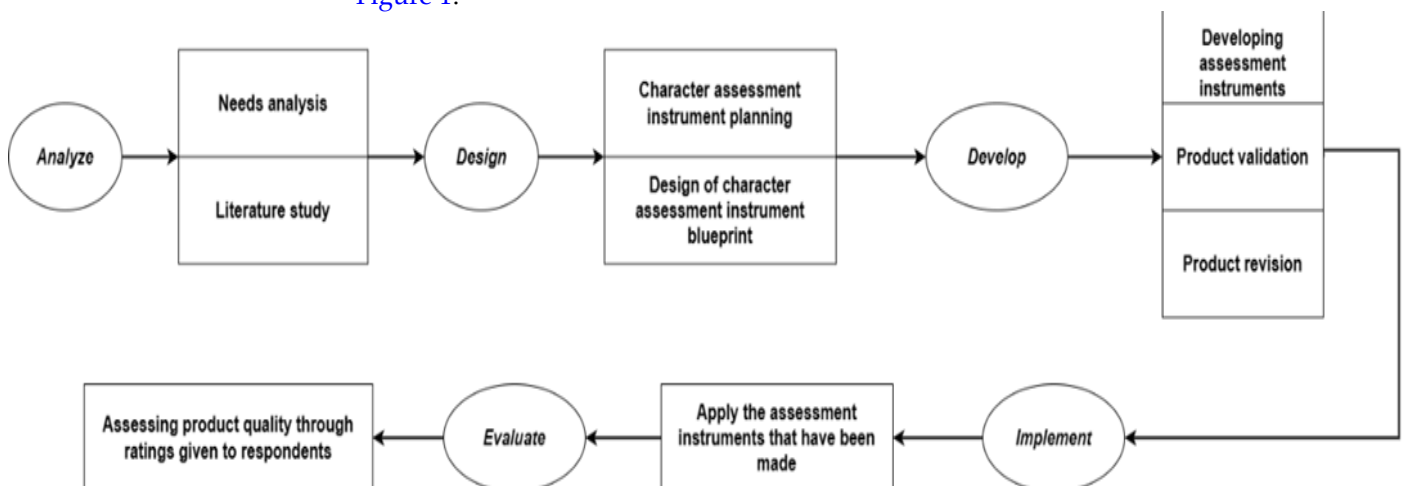

Figure 1. Flow of research stages combination and modification of ADDIE

2.1 Analyze

The initial stage within the ADDIE framework involves analysis, where emphasis is placed on needs assessment, aiming to identify the characteristics of learners, ascertain the behavioral outcomes of learners, and determine whether the core values of the nation are integrated into English language instruction (Pulukadang et al., 2023). Subsequently, a detailed analysis, encompassing needs and task analysis, is conducted based on the necessity to investigate and elucidate whether the encountered challenges require resolution through the development of character evaluation instruments that are more precise, diverse, valid, and reliable in assessing student character (Idris & Rus, 2023). The analytical activities are carried out through observation and interview interactions, to observe the

realities in the field. This stage encompasses the establishment of instructional objectives and the formulation of task concepts to be instructed, ultimately contributing to the design phase. Each stage will be explained in more detail in the results section.

2.2 Design

The design phase aims to create the framework or blueprint of the assessment instrument for students' character (Feng & Sangsawang, 2023). This process commences with planning and progresses to designing, involving the determination of instrument specifications, establishment of the instrument scale, scoring system determination, formulation of instrument indicators-items, and instrument item composition. Determination of instrument specifications encompasses the development of a valid and reliable assessment tool with a grid based on the conceptual definition from international literature. Learning theories serve as the foundation for grid development. Clear specifications aid in obtaining accurate assessment results. The subsequent stage involves composing instruments with questions aligned to specifications. The instruments are closed-ended, utilizing a Likert scale with 4 response options. The selection of the Likert scale is based on its ease of use and reliability. Gradation determination is adjusted to the emotional and social aspects of student development. The scoring system facilitates result interpretation with a Likert scale ranging from strongly agree to strongly disagree. The highest score is 4, while the lowest is 1. Formulating instrument indicators-items involves understanding goals, determining competencies, and composing detailed indicator items. Designing instrument items begins with a clear and objective understanding of assessment goals, with testing and evaluating the instrument's effectiveness and reliability (Yoong, 2022).

2.3 Develop

The developmental stage serves as a pivotal step in transforming the prototype design into an operational and tangible product, as elucidated by Nurmiati et al. (2023). Within this phase, designers amalgamate and incorporate crucial materials generated during the design stage, encompassing the creation of content, media, and activities. Subsequently, rigorous testing of the materials is conducted to assess their usability, functionality, and overall effectiveness, as detailed by Shakeel et al. (2023). The product development phase involves four key steps to achieve the final form of the student character evaluation model after revision based on input and competency test data.

Firstly, expert validation is conducted to obtain input from designated experts. Experts review questions, language, and grammar in the instrument, ensuring diversity and reader interest. The validation results by experts serve as the basis for improvements in the development of student character assessment instruments. The second step is instrument testing, involving empirical testing on participants. The initial test was conducted on 331 junior high school students in Yogyakarta to assess construct validity using exploratory factor analysis (EFA). The second test on 210 students was used to confirm hypotheses and perform confirmatory factor analysis (CFA) to validate the model from EFA results.

The study tested an instrument designed for character assessment on secondary school students in the Yogyakarta province, who were implementing Curriculum 2013 during the academic year 2022/2023. The selection of schools involved purposive sampling, focusing on state junior high schools in Yogyakarta based on new student admissions in the city of Yogyakarta. The research covered six public and one private junior high school in Yogyakarta, with a specific emphasis on schools with the highest and lowest interest values or ranks according to the statistics for first-choice applicants in 2023. Private schools were selected based on their religious affiliation in the city. This study employs various data collection techniques and instruments to assess character development in English education. Structured interviews with English teachers are used for need analysis, while unstructured interviews gauge the effectiveness of a character assessment

questionnaire during the evaluation stage. The research also involves self-assessment tools and questionnaires administered to education and material experts for data collection.

Next, the third step involves analyzing the instrument to obtain estimates of the developed instrument's characteristics. This analysis focuses on the scale model used, examining differential item functioning and instrument reliability. Construct validity and instrument reliability are tested using both EFA and CFA. Data analysis includes content validity using the Aiken formula, construct validity through EFA and CFA, and reliability estimation using Cronbach's Alpha. The effectiveness of the instrument is evaluated based on the implementation process and feedback from participants who have utilized the character assessment tool. Finally, after testing and analyzing the instrument, product revision is undertaken. Instrument items that do not meet criteria are improved to enhance the quality, performance, and effectiveness of the instrument. The revised instrument is expected to measure student character and meet the criteria for a well-qualified instrument.

2.4 Implementation

The revised product is implemented through the process of measurement, wherein the outcomes of the measurements are interpreted. The steps involved in this process are as follows. Firstly, conducting measurements or tests after completing the instrument development phase and undertaking post-pilot revisions. The developed test instruments are then administered to students as respondents to be answered according to the established schedule. Monitoring or supervision is necessary to ensure the integrity of the test implementation by the learners. According to Nurmiati et al. (2023), implementation involves the tangible enactment of deploying or executing products that have been deemed sound subsequent to undergoing the validation procedure through the process of measurement.

The results of these measurements are manifested in the form of numerical values derived from the accumulation of participants' scores for each presented statement. Secondly, interpreting the measurement results represented in the form of scores or numbers. This interpretative process is known as assessment and requires a set of criteria depending on the scale and number of items applied. Assessment is conducted based on four categories, with the final assessment referring to the total score of measurement outcomes compared to the criteria for the attainment of participants' character.

2.5 Evaluate

The evaluation phase constitutes a stage involving the analysis of assessment instruments, conducted at each conclusion of the research phases, spanning from analysis, design, development, to implementation. Within this evaluative stage, an analysis of the obtained data is performed to assess the extent to which the developed product can be deemed feasible, valid, and reliable.

3. Results

The outcome of this investigation encompasses the creation of a student character assessment instrument tailored for junior high school levels, aligned with the 2013 Curriculum, and implemented across junior high schools in the Yogyakarta city region. This tool has been meticulously devised to facilitate educators in conducting precise and thorough evaluations of students' character within the context of English learning. Its utilization enables the acquisition of more precise data, facilitating effective analysis and comprehension of students' character development. Furthermore, this evaluative instrument serves as a strategic resource for formulating pertinent learning programs, thereby contributing positively to the enhancement of teaching quality and the attainment of desired student learning outcomes. The development process of this character

assessment instrument adheres to the ADDIE method, encompassing five sequential stages: Analysis, Design, Develop, Implementation, and Evaluation. These stages are elucidated in detail within the research findings.

3.1 Analysis

This section outlines the findings of a study on the analysis and design of a character assessment instrument. The initial research results on product development are shared, underscoring the importance of teacher involvement and a thorough literature review. The objective is to position the character assessment instrument as a strategic tool for planning effective learning programs, ultimately enhancing teaching quality and student learning outcomes. The integrated character assessment in English language learning, facilitated by self-assessment, aims to nurture positive and resilient character traits, promoting self-awareness, personal responsibility, and proactive planning for behavioral and attitudinal improvements.

3.2 Design

The study designed character assessment instruments, identifying five core character values: Religious, nationalist, independent, cooperative, and integrity. These values are central to the assessment, comprising a total of 12 indicators and 50 items in the resulting statement. The developed assessment tool takes the form of a sheet for individuals to fill out, encouraging reflection on the incorporation of these character values into their daily lives. The study successfully produced compelling results for measuring character development in individuals or groups within an educational context. Interviews with teachers revealed concerted efforts to integrate character values into every classroom activity, indicating a commitment to improving learning effectiveness.

3.3 Develop

In this section, the primary objective is the formulation of a robust and reliable model for assessing student character. Prior to subjecting the assessment instrument to respondent data, content validation procedures are employed to ascertain its alignment with research objectives. A panel of five validators, proficient in psychometrics and language, validates the content of the self-assessment instrument designed for appraising students' scientific thinking skills, comprising a set of 50 items.

Following the content validation phase, the initial trial encompasses a cohort of 331 learners serving as respondents, with data analysis conducted through Exploratory Factor Analysis (EFA) utilizing SPSS. Items exhibiting a component matrix loading below 0.400 are deemed inadequate and subsequently excluded. The EFA reveals the identification of 7 factors associated with student character, leading to the reduction of the initial 50 items to a refined set of 24 items. Items are eliminated based on factor loadings falling below 0.400, while loadings exceeding this threshold contribute to the identification of pivotal factors associated with the loaded items. This systematic process serves to enhance the precision of the instrument, thereby ensuring its efficacy in effectively measuring students' character (Mohamed et al., 2021). Moreover, the determination of factor discoveries generated within the tool can be discerned via scree-plot examination, illustrated in the visual depiction presented in Figure 2.

The eigenvalue distribution graph exhibits a discernible slope pattern indicative of the intensity of various factors. Notably, the eigenvalue begins to slope from the eighth factor onward, suggesting the prominence of seven dominant factors within the English character assessment instrument designed for junior high school students. These seven factors collectively account for 62.99% of the variance in the measurement outcomes. Subsequently, Confirmatory Factor Analysis (CFA) was conducted in the second phase, employing a two-level latent variable model through RStudio software to validate and affirm the previously proposed factor model for measuring the targeted construct or variable. The analysis utilized empirical data derived from the earlier Exploratory Factor Analysis (EFA).
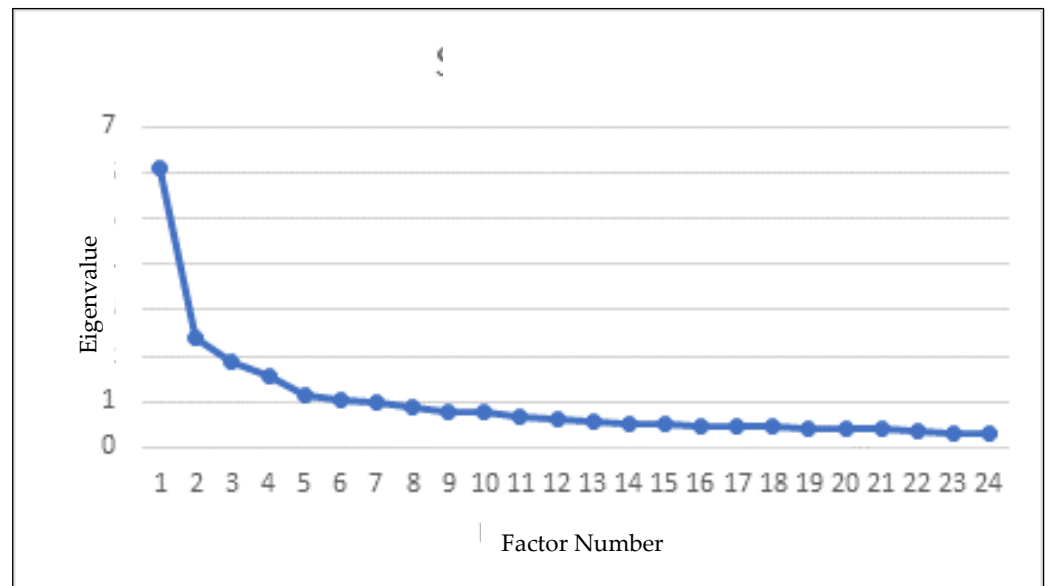
Figure 2. Scree plot of exploratory factor analysis results

The second trial, involving 210 students from two schools in Yogyakarta, employed CFA results analyzed in RStudio to establish construct validity. The analysis revealed seven factors: integrity, resilience, cooperation, independence, perseverance, tolerance, and collaboration. However, three items were identified with loading factors below 0.4, indicating their invalidity. Consequently, these three invalid items and an entire factor were excluded from the analysis, resulting in a revised test with six factors and a total of 21 items. Despite these adjustments, some items still exhibited loading factors below 0.4, and the model fit analysis, employing the second-order model, indicated inadequate fit.

The results underscored the necessity for adjustments and refinements to the factor model, a crucial undertaking to ensure its congruence with empirical observations. This pivotal process underscores the significance of enhancing the model to conform to the criteria for optimal model fit during the construction of measurement instruments or the evaluation of theories. Successful model fitting is contingent upon the loading factor in Confirmatory Factor Analysis (CFA) meeting the stipulated threshold of > 0.4 (Retnawati, 2016). Moreover, enhancing model adequacy can be achieved through the introduction of error covariance among factors or items (Suryani, 2014). In summary, the developed model underwent modifications from its original hypothesized form. The adjustments involved removing items or factors with factor loadings below 0.4 and introducing error covariance between factors, indicating relationships contributing to measurement errors. Following these modifications, Confirmatory Factor Analysis (CFA) resulted in a construct validity testing revealing six factors: integrity, resilience, cooperation, independence, perseverance, and tolerance. The construct validation, supported by the loading factors calculated, is visually represented in Figure 3. Figure 3 demonstrates the path diagram, the relationships among these factors, and their indicators comprising 19 items.

The modified test results, as illustrated in Figure 3, reveal that all factor loading values for each factor and item surpass the threshold of 0.4. The model's adequacy is supported by statistical evidence: the p-values obtained are 0.056 (slightly greater than 0.05). Additionally, key fit indices confirm the suitability of the model—Root Mean Square Error of Approximation (RMSEA) is 0.031 (below 0.05), Comparative Fit Index (CFI) is 0.979 (exceeds 0.9), Tucker-Lewis Index (TLI) is 0.972 (exceeds 0.9), and Standardized Root Mean Square Residual (SRMR) is 0.055 (below 0.08). The second-order model approach supports the entire model fitting well. Finally, the reliability test results for each dimension or factor, post-modification, are detailed in Table 1.
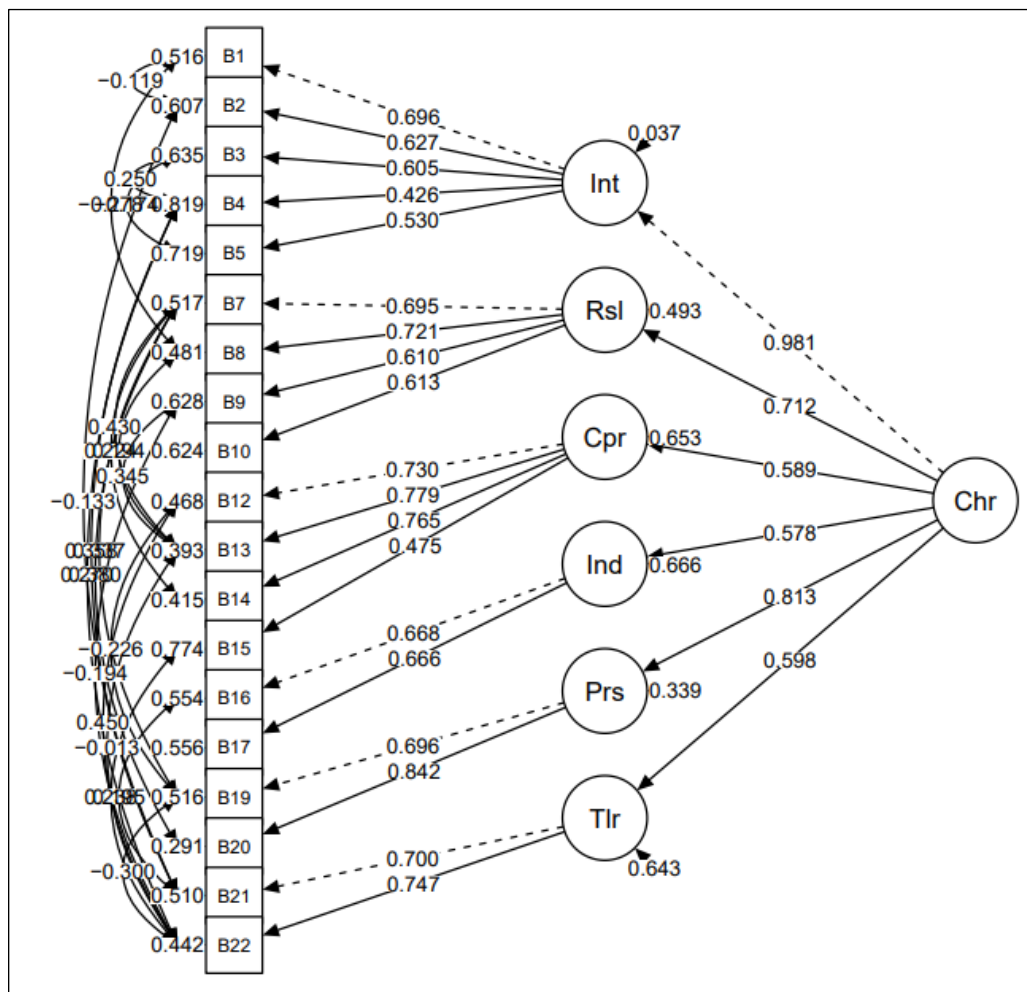
Figure 3. Confirmatory factor analysis model of character assessment

Table 1. Reliability results

Factor	Item	Factor Name	Cronbach Alpha	Criteria
1	B1; B2; B3; B4 & B5	Integrity	0,715	Reliabel
2	B7; B8; B9 & B10	Resilient	0,765	Reliabel
3	B12; B13; B14 & B15	Cooperation	0,768	Reliabel
4	B16 & B17	Independent	0,604	Unreliable
5	B19 & B20	Perseverance	0,718	Reliabel
6	B21 & B22	Tolerance	0,715	Reliabel

The table indicates good scores, with reliability values ranging from 0.604 to 0.768 for each dimension or factor evaluated. However, one factor, identified as factor 4 (independent), exhibited a Cronbach Alpha value below 0.7, signifying unreliability. Consequently, factor 4, consisting of 2 items (B16 and B17), was excluded from the analysis. The revised analysis now comprises five factors and 17 items, all demonstrating good validity and reliability based on expert validation, as well as the results of EFA, CFA, and the reliability tests conducted between factors.

3.4 Implementation

The section discusses the analysis of the character assessment instrument's measurement results. It starts with a thorough examination of the final study product, providing a detailed insight into the instrument's characteristics and quality. The assessment comprised self-assessment by 85 students, yielding data based on responses to 17 items rated on a scale of 1 to 4. Scores ranged from 37 to 68, with an optimal average

of 52.5 and a standard deviation of 5.17. A visual representation illustrates the percentage distribution of learner character categories in Figure 4.
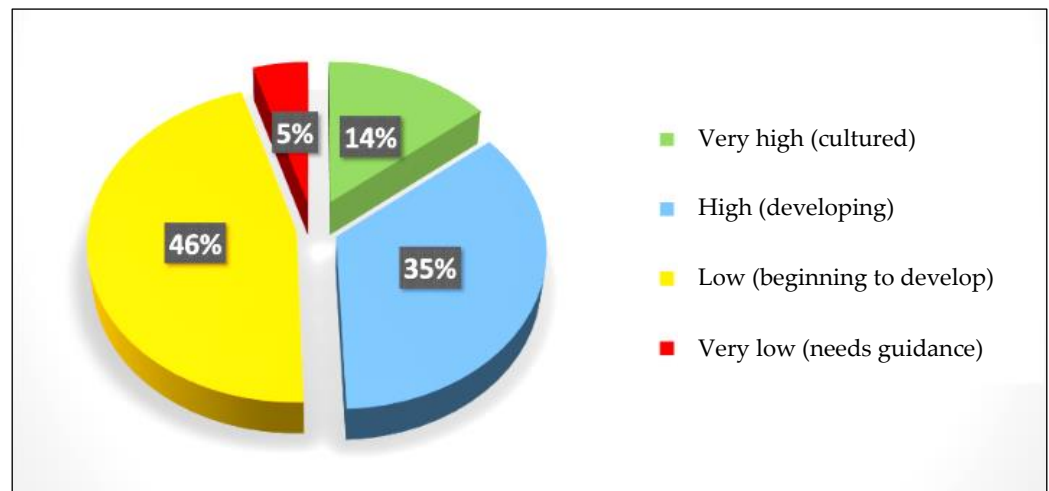

Figure 4. Percentage of learner character categories

The pie chart analysis of surveyed students indicates that 14% exhibit a very high/cultured character, 35% fall into the high/developing category, 46% are in the low/starting to develop category, and 5% are classified as very low/requiring guidance. Notably, the low/starting to develop category has the highest percentage, suggesting a substantial number of students fall into this range. Conversely, the very low/requires guidance category has the lowest percentage. This underscores the instrument's significance, emphasizing the need for accuracy and reliability during evaluation to ensure it consistently measures the intended characteristics of students.

3.5 Evaluation

The evaluation procedure in the evaluation phase of the research will use the general procedure proposed by Branch (2009), which involves establishing evaluation criteria, selecting tools, and conducting assessments. According to Branch (2009), the objective of the evaluation stage is to appraise the caliber of instructional artifacts and procedures, both antecedent to and after execution. This evaluative procedure encompasses a sequence of pivotal measures aimed at discerning deficiencies, enhancing, and refining the apparatus before its formal utilization.

In the study, unstructured interviews with seven students from SMP Negeri 15 Yogyakarta, who utilized the character assessment instrument, revealed positive feedback. The summarized evaluation results are as follows: (1) The character assessment instrument, still under development, exhibits potential for effectively evaluating student character; (2) The same instrument consistently demonstrates its potential for assessing student character effectively; (3) The current character assessment instrument has reached an appropriate developmental stage; (4) The instrument shows potential for widespread application among students and adoption by various schools. Based on these findings, the study establishes a solid foundation for reflection and improvement in developing a more effective character education program. Further evaluation of the instrument confirms its effective potential in assessing and evaluating student character.

4. Discussion

4.1 Analysis

In this stage, a structured interview was conducted with every English teacher in five junior high schools in Yogyakarta. The interviews were aimed at finding out the planning, implementation, and assessment of character education used in schools. During the interviews, it was discovered that English teachers had identified all the character values that can be incorporated into their lessons. They already have a list of character

values that are integrated into each English subject and match them with the syllabus. They choose specific characters that will be integrated into certain chapters or lessons so that students can benefit from and achieve the goals of integrated character education.

This finding is in line with Sumague's (2020) view, which states that a qualified language educator should combine several attributes to effectively convey knowledge to students. Although it is a challenge for English teachers to embed character values into learning activities and design effective learning (Sofiana et al., 2019). To enhance English learning comprehensively, it is crucial to thoughtfully assess learning outcomes, including evaluating students' character. An organized educational process that addresses all facets of learning plays a vital role in improving overall efficiency in English learning. (Malaikosa & Taopan, 2020).   Then in terms of the implementation of the integration of the cultivation of character values in English subjects in the classroom, namely getting students used to praying before and after completing learning activities. Providing character values that have been integrated with learning materials, providing examples of characters that children can emulate. In the implementation of the integration of the cultivation of character values in English subjects, there are also inhibiting factors, namely the background of students, and the environment of the students, which must always repeat to give advice or repeat the invitation to do good to the students.

This indirectly provides support for the theoretical foundations proposed by   Darnell et al. (2019); Dariyanto and Awiria (2023). The study explores the incorporation of moral knowledge into the learning process. Teacher interviews reveal that successful education integration shapes beliefs and influences students' character development. Character values, subtly embedded in learning activities rather than explicitly taught, contribute to positive character formation. This approach aligns with targeted program goals tailored to different stages of child development. In essence, students with a deep understanding of moral knowledge are likely to intellectually comprehend and emotionally appreciate these values, fostering the growth of moral sentiments. The ultimate objective is to mold children's personalities, enabling them to become quality, responsible, and ethical citizens capable of fostering positive interactions in society and the state (Afifuddin & Burga, 2022; Jabbarova, 2022). Based on the outcomes of this study, the developed character assessment instrument holds the potential for contributing to character education in Indonesia.

4.2 Design

Researchers have designed questions and statements to assess students' character towards English subjects. The first step in creating a tool to measure student character in English learning through self-assessment is the planning phase, which involves two steps: general planning and specific design for the self-assessment tool. The planning process includes designing the character assessment approach, drafting the character assessment tool, and producing an initial draft of the character assessment tool that has a scientific foundation.

Following what has been theorized by Feng and Sangsawang (2023), the design stage aims to create a framework or blueprint for student character assessment instruments. The first step in designing the character assessment instrument in this study is to identify the core values to be assessed. In this study, the main focus is on the five core values of national character: religion, nationalism, independence, cooperation, and integrity. A draft character assessment instrument is prepared after identifying the main purpose, and core values, and designing the assessment. It consists of assessment grids, criteria, and learner assessment sheets. Furthermore, these character aspects were developed into 12 indicators, and the total number of questions or statements produced was 50 items. This study emphasizes self-assessment and has designed an assessment sheet that individuals can fill in to measure the extent to which they practice certain character values in their daily lives.

4.3 Development

This development research has provided evidence to support the theory (Vogt, 2018). Regarding the validity of data collection tools (instruments), which is the most complicated and most important thing to consider when conducting or criticizing a study. This can be seen from how the stages of the research instrument are validated starting from validation by experts (Retnawati, 2016; Azwar, 2021; Istiyono, 2020), then construct validation to explore that the proposed factor model matches the empirical data obtained from the Exploratory Factor Analysis (EFA) analysis (Yong & Pearce, 2013; Retnawati, 2016; Watkins, 2018; Hair et al., 2019; Mohamed et al., 2021; Steenkamp & Olivares, 2023). Then to prove that the proposed factor model is appropriate, Confirmatory Factor Analysis (CFA) is applied to confirm the empirical data obtained from Exploratory Factor Analysis (EFA) (Yong & Pearce, 2013; Brown, 2015; Retnawati, 2016).

However in this study, going through the approach through CFA to test how well the data fit or not with the proposed model or theory is sufficient (Retnawati, 2016; Brauer et al., 2023; Panahi et al., 2023). To ensure valid and reliable research instruments, this study successfully developed instruments according to the theory proposed by Amalia et al. (2023). This can be seen through construct validity and reliability tests that produce five factors and 17 statements that have good validity and reliability.

4.4 Implementation

During the implementation phase, the developed instrument is subjected to real testing situations to assess its functionality and effectiveness. The testing process involves administering the instrument to a representative sample group, which may include individuals from the target audience or those who closely resemble the intended participants. The focus during this phase is not only on the accuracy of the instrument in measuring the desired character traits but also on its practicality and user-friendliness. Participants' feedback and observations are gathered to identify any potential issues related to the clarity of instructions, comprehensibility of questions, and overall usability of the instrument.

Moreover, the real testing situation allows for the identification of unforeseen challenges and provides an opportunity to make necessary adjustments to enhance the instrument's performance. This implementation stage has resulted in the suitability of the results obtained as a result of the application of the theory by Nurmiati et al. (2023) which states implementation stage involves the tangible enactment of deploying or executing products that have been deemed sound after undergoing the validation procedure through the process of measurement. This iterative process of testing, collecting feedback, and refining the instrument is crucial for ensuring that the final version meets the standards of validity and reliability while being well-received by the intended users. Ultimately, the implementation phase serves as a valuable step in refining and optimizing the developed instrument before its broader application in character assessment contexts.

4.5 Evaluate

The evaluation procedure at the evaluation stage in the research will use the general procedure proposed by Branch (2009) which consists of determining evaluation criteria, selecting evaluation tools, and conducting evaluations. Evaluation is conducted to assess the effectiveness of the character assessment instrument. The validity, reliability, and sustainability of the instruments are evaluated to ensure that they can provide accurate and reliable data in measuring character development.

Furthermore, based on the findings of this study, there is evidence to support the concept promoted by Istiyono (2020), which asserts that the assessment system implemented in each educational institution should have the ability to: 1) provide accurate information; 2) encourage students to learn; 3) motivate teachers to teach; 4) improve institutional performance; 5) improve the quality of education. This can be seen from the character assessment instrument integrated into English learning for junior high school students has been proven in terms of good validity and reliability, then based on the

evaluation results regarding the assessment instrument in the form of self-assessment that has been implemented shows that students are encouraged to be even better at evaluating character after assessing themselves using the developed self-assessment assessment instrument. This assessment instrument is not only intended for general use in various schools but can also be a guideline in designing more effective character assessment instruments in the context of English language learning for teachers, schools, and researchers. In addition, this instrument can also be a guide for other educators and researchers who want to develop student character assessment instruments at various levels of education.

5. Conclusions

This research successfully devised and authenticated a character assessment tool specifically crafted for gauging five fundamental dimensions of national character in student populations, encompassing religious, nationalist, independent, cooperative, and integrity values. The meticulous developmental procedure encompassed content validation, Exploratory Factor Analysis (EFA), and Confirmatory Factor Analysis (CFA), yielding an enhanced instrument featuring six factors and seventeen dependable items. The study's outcomes furnish a significant evaluative instrument for character progression in students, underpinned by a comprehensive validation process, rendering the tool robust and applicable. This instrument stands as an asset for educators, researchers, and policymakers to evaluate and foster character development, thereby contributing to the enhancement of the educational framework and the cultivation of well-rounded individuals.

Author Contributions: Methodology, collecting and analysis data, writing—original draft preparation, Manggaberani, A.A.; review, validation and editing, Putro, N.H.P.S.

Conflicts of Interest: Authors declare there are no conflicts of interest.

6. References

Afifuddin, A., & Burga, M. A. (2022). Religious ethics education of muslim and buddhist families in Watampone. *Al-Qalam Jurnal Penelitian Agama dan Sosial Budaya*, *28*(2), 228–238. https://doi.org/10.31969/alq.v28i2.1098

Akhter, S., Haidov, R., Majeed Rana, A., & Hussain Qureshi, A. (2020). Exploring the significance of speaking skill for EFL learners. *PalArch's Journal of Archaeology of Egypt / Egyptology*, *17*(9), 6019–6030. https://archives.palarch.nl/index.php/jae/article/view/5149

Alivi, F. J., & Widihastuti. (2022). Development of affective self-assessment instrument of chemistry for high school student as the daily assessment guideline. *European Journal of Educational Research*, *11*(1), 445–456. https://doi.org/https://doi.org/10.12973/eu-jer.11.1.445

Amalia, A. N., Suyono, S., & Arthur, R. (2023). *Penyusunan instrumen penelitian: Konsep, teknik, uji validitas, uji reliabilitas, dan contoh instrumen penelitian* (1st ed.). PT Nasya Expanding Management. https://books.google.co.id/books/about/Penyusunan_Instrumen_Penelitian.html?id=rx3JEAAAQBAJ&redir_esc=y

Arif, A. M., Nurdin, N., & Elya, E. (2023). Character education management at islamic grassroot education: The integration of local social and wisdom values. *Al-Tanzim : Jurnal Manajemen Pendidikan Islam*, *07*(02), 435–450. https://doi.org/10.33650/al-tanzim.v7i2.5468

Aydin, A., Gürsoy, A., & Karal, H. (2023). Mobile care app development process: using the ADDIE model to manage symptoms after breast cancer surgery (step 1). *Discover Oncology*, *14*(1). https://doi.org/10.1007/s12672-023-00676-5

Azwar, S. (2021). *Reliabilitas dan validitas* (4th ed.). Pustaka Pelajar. https://inlislite.uin-suska.ac.id/opac/detail-opac?id=1829

Branch, R. M. (2009). *Instructional design: The ADDIE approach*. Springer. https://link.springer.com/book/10.1007/978-0-387-09506-6

Brauer, K., Ranger, J., & Ziegler, M. (2023). Editorial: Confirmatory factor analyses in psychological test adaptation and development: A nontechnical discussion of the WLSMV estimator. *Psychological Test Adaptation and Development*, *4*(1), 4–12. https://doi.org/10.1027/2698-1866/a000034

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). Guilford Press. https://www.guilford.com/books/Confirmatory-Factor-Analysis-for-Applied-Research/Timothy-Brown/9781462515363

Dariyanto, D., & Awiria, A. (2023). Pendidikan karakter pada pembelajaran percakapan Bahasa Inggris siswa sekolah dasar berasrama. *Educational Journal of Bhayangkara (EDUKARYA)*, *3*(1), 65–74. https://doi.org/10.31599/edukarya.v3i1.2534

Darnell, C., Gulliford, L., Kristjánsson, K., & Paris, P. (2019). Phronesis and the knowledge-action gap in moral psychology and moral education: A new synthesis? *Human Development*, *62*(3), 101–129. https://doi.org/10.1159/000496136

Feng, J., & Sangsawang, T. (2023). Information technology, according to the ADDIE model on English subject teaching, enhances the learning achievement of Shunde Polytechnic students in China. *TOJET: The Turkish Online Journal of Educational Technology*, *22*(4), 121–131. https://files.eric.ed.gov/fulltext/EJ1409897.pdf

García-Moriyón, F., González-Lamas, J., Botella, J., Vela, J. G., Miranda-Alonso, T., Palacios, A., & Robles-Loro, R. (2020). Research in moral education: The contribution of P4C to the moral growth of students. *Education Sciences*, *10*(4). https://doi.org/10.3390/educsci10040119

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage. https://books.google.co.id/books/about/Multivariate_Data_Analysis.html?id=0R9ZswEACAAJ&redir_esc=y

Harmanto, B., Anwar, K., Arifani, Y., Basri, H., & Nuruddin, M. (2023). The relationship between attitudes, motivations and gender in learners of english for academic purposes. *International Journal of Instruction*, *16*(3), 971–990. https://doi.org/10.29333/iji.2023.16352a

Heale, R., & Twycross, A. (2015). Validity and reliability in quantitative studies. *Evidence-Based Nursing*, *18*(3), 66–67. https://doi.org/10.1136/eb-2015-102129

Heritage, M., Walqui, A., & Linquanti, R. (2020). *English language learners and the new standards: Developing language, content knowledge, and analytical practices in the classroom*. Harvard Education Press. https://hep.gse.harvard.edu/9781612508016/english-language-learners-and-the-new-standards/

Hidayati, K., Budiyono, budiyono, & Sugiman, sugiman. (2018). Development and validation of student's responsibility scale on mathematics learning using subject scaling model. *International Journal of Instruction*, *11*(4), 499–512. https://doi.org/10.12973/iji.2018.11431a

Ibrahim, F. M., Abdelaziz, E., Dabou, R., Abdelsamad, S., & Abuijlan, I. A. M. (2024). Prevalence of bullying and its impact on self-esteem , anxiety and depression among medical and health sciences university students in RAS Al Khaimah , UAE. *Heliyon*, *10*(3), e25063. https://doi.org/10.1016/j.heliyon.2024.e25063

Idris, M. R., & Rus, R. C. (2023). Developing the rekatronik module using ADDIE model for electronic design topic in the design and technology (D&T) subject. *Asian Journal of Research in Education and Social Sciences*, *5*(2), 113–122. https://doi.org/10.55057/ajress.2023.5.2.13

Istiyono, E. (2020). *Pengembangan instrumen penilaian dan analisis hasil belajar fisika dengan teori tes klasik dan modern* (2nd ed.). UNY Press. http://library.uny.ac.id/sirkulasi/index.php?p=show_detail&id=57737

Jabbarova, M. K. (2022). The importance of education directed to the personality of the child in ensuring the education, upbringing, physical, mental and spiritual maturity of the child. *Models and Methods for Increasing the Efficiency of Innovative Research*, *2*(17), 54–59. https://interonconf.org/index.php/ger/article/view/100

Jurianto, J. (2017). Model pengembangan desain instruksional dalam penyusunan modul pendidikan pemustaka (Library Instruction). *Jurnal Media Pustakawan*, *24*(3), 32–39. https://doi.org/https://doi.org/10.37014/medpus.v24i3.461

Kleemann, S. S., Franikowski, P., & Nicolai, S. (2023). Development and validation of a scale to assess moral disengagement in high-carbon behavior. *Sustainability*, *15*(3), 1–23. https://doi.org/10.3390/su15032054

Koenig, E., Guertler, K., Ĭarnowska, D., & Horbaƥauskieno, J. (2020). Developing English language competence for global engineers. *IEEE Global Engineering Education Conference, EDUCON*, 242–249. https://doi.org/10.1109/educon45650.2020.9125208

KPAI R.N. (2021). *Data kasus perlindungan anak 2016 – 2020*. https://bankdata.kpai.go.id/tabulasi-data/data-kasus-perlindungan-anak-2016-2020

Liando, N. V. F., & Tatipang, D. P. (2022). English or Indonesian language? parents' perception toward children's second language learning context. *Jurnal Lingua Idea*, *13*(1), 61–75. https://doi.org/10.20884/1.jli.2022.13.1.5749

Malaikosa, C. A., & Taopan, L. L. (2020). Teaching English at junior high school in Indonesian rural area: The implementation of scientific approach. *Lensa: Kajian Kebahasaan, Kesusastraan, dan Budaya*, *10*(2), 206. https://doi.org/10.26714/lensa.10.2.2020.206-217

Mardapi, D. (2017). *Pengukuran, penilaian, dan evaluasi pendidikan* (2nd ed.). Parama Publishing. https://opac.perpusnas.go.id/DetailOpac.aspx?id=1162325

Marela, G., Wahab, A., & Marchira, C. R. (2017). Bullying verbal menyebabkan depresi pada remaja SMA di kota Yogyakarta. *Berita Kedokteran Masyarakat*, *33*(1), 43–48. https://doi.org/doi.org/10.22146/bkm.8183

Mohamed, A., Abdullah, Z., & Zabidi Abdul Razak, A. (2021). Validation of an instrument for measuring integrated principal leadership practices. *The Malaysian Online Journal of Educational Management (MOJEM)*, *9*(1), 1–20. https://ejournal.um.edu.my/index.php/MOJEM/article/view/27958

Nisa, K., Nurhasanah, N., Kusuma, A. S. H. M., Deni Sutisna, D. S., & Sari, M. P. (2021). Pengembangan instrumen penilaian karakter untuk siswa sekolah dasar (SD). *Jurnal Ilmiah Profesi Pendidikan*, *6*(1), 102–107. https://doi.org/10.29303/jipp.v6i1.158

Nurmiati, N., Arismunandar, A., & Rauf, B. A. (2023). Problem-based learning based biology learning model development using the environment to improve motivation and learning outcomes of biology high school students. *Asian Journal of Education and Social Studies*, *49*(2), 104–112. https://doi.org/10.9734/AJESS/2023/v49i21123

Panahi, S., Bazrafshani, A., & Mirzaie, A. (2023). Development and validation of a modified LibQUAL scale in health sciences libraries: Application of structural equation modeling. *Journal of the Medical Library Association*, *111*(4), 792–801. https://doi.org/10.5195/jmla.2023.1348

Pulukadang, M. A., Baruadi, M. K., Malabar, S., & Hulukati, E. P. (2023). Development of cultural insights courses (Characterizing of Ung) based on ADDIE approach in the faculty of literature and culture Universitas Negeri Gorontalo. *International Journal of Education and Social Science Research*, *6*(1), 232–244. https://doi.org/10.37500/ijessr.2023.6116

Rao, P. S. (2019). The role of English as global language. *Research Journal of English (RJOE)*, *4*(1), 65–79. https://www.researchgate.net/publication/334282978_the_role_of_english_as_a_global_language#fulltextfilecontent

Retnawati, H. (2016). *Validitas, reliabilitas, dan karakteristik butir*. Parama Publishing. http://library.fip.uny.ac.id/opac/index.php?p=show_detail&id=9172

Sadri, M., Adisaputera, A., & Daulay, S. (2020). The development of character-based assessment instruments in poetry writing material for class X students of SMA Negeri 3 Medan. *Budapest International Research and Critics in Linguistics and Education (BirLE) Journal*, *3*(4), 1884–1905. https://doi.org/10.33258/birle.v3i4.1411

Shakeel, S. I., Al Mamun, M. A., & Haolader, M. F. A. (2022). Instructional design with ADDIE and rapid prototyping for blended learning: validation and its acceptance in the context of TVET Bangladesh. *Education and Information Technologies*, 1–30. https://doi.org/10.1007/S10639-022-11471-0

Sofiana, N., Mubarok, H., & Yuliasri, I. (2019). English language teaching in secondary schools: An analysis of the implementation of Indonesian ELT 2013 curriculum. *International Journal of Instruction*, *12*(1), 1533–1544. https://doi.org/10.29333/iji.2019.12198a

Steenkamp, J. B. E. M., & Olivares, A. M. (2023). Unrestricted factor analysis: A powerful alternative to confirmatory factor analysis. *Journal of the Academy of Marketing Science*, *51*(1), 86–113. https://doi.org/10.1007/s11747-022-00888-1

Sumague, J. A. (2019). *Fundamental concepts of language teaching* (2nd ed.). Society Publishing. https://www.amazon.com/Fundamental-Concepts-Language-Teaching-Julieta/dp/1774072661

Suryani. (2014). Persepsi nilai yang diterima nasabah pada layanan perbankan Syariah: Analisis konfirmatori faktor. *Finance and Banking Journal*, *16*(1), 104–124. https://journal.perbanas.id/index.php/jkp/article/view/204/31

Susanto, S., Ritonga, A. W., & Desrani, A. (2022). The challenge of the integrated character education paradigm with 21st-century skills during the Covid-19 pandemic. *Cendekia: Jurnal Kependidikan dan Kemasyarakatan*, *1*(1), 85–104. https://doi.org/10.21154/cendekia.v1i1.3816

Tahmasbi, S., Hashemifardnia, A., & Namaziandost, E. (2019). Standard English or world Englishes: Issues of ownership and preference. *Journal of Teaching English Language Studies*, *7*(3), 83–98. https://journals.iau.ir/article_665009_bc918242d35f565d74a0b19073056696.pdf

Tirrell, J. M. (2022). Forgiveness as a character strength: Toward a developmental model and research agenda. *Journal of Moral Education*, *51*(3), 312–335. https://doi.org/10.1080/03057240.2021.1873754

Vogt, K. (2018). Assessment—what, why and how? In C. Surkamp & B. Viebrock (Eds.), *Teaching English as a foreign language an introduction* (p. 251). J. B. Metzler Verlag. https://doi.org/10.1007/978-3-476-04480-8_14

Watkins, M. W. (2018). Exploratory factor analysis: A guide to best practice. *Journal of Black Psychology*, *44*(3), 219–246. https://doi.org/10.1177/0095798418771807

Wijayanti, D., & Pratomo, W. (2019). Pendidikan karakter melalui model pembelajaran kreatif bagi siswa sekolah dasar (Studi di SDN Mendungan 2 Yogyakarta). *Taman Cendekia: Jurnal Pendidikan Ke-SD-An*, *3*(1), 276–282. https://doi.org/https://doi.org/10.30738/tc.v3i1.4291

Yong, A. G., & Pearce, S. (2013). A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *The Quantitative Methods for Psychology*, *9*(2), 79–94. https://doi.org/https://doi.org/10.20982/tqmp.09.2.p079

Yoong, S. M. (2022). Applying the ADDIE model to design and develop an instrument for dyscalculia. *Seameo*, *1*, 55–68. https://www.researchgate.net/publication/366861346_Applying_the_ADDIE_Model_to_Design_and_Develop_an_Instrument_for_Dyscalculia

Yu, S., & Setiyaningrum, A. (2018). Studi mengenai city branding kota Yogyakarta sebagai kota pelajar di Indonesia. *Matrik: Jurnal Manajemen, Strategi Bisnis dan Kewirausahaan*, *2*(1), 31–46. https://doi.org/10.24843/matrik:jmbk.2019.v13.i01.p04

Yusoff, A. M., Madihie, A., & Hutasuhut, I. J. (2023). Efikasi Kendiri sebagai moderator antara metakognitif dan resilien dalam kalangan guru bimbingan dan kaunseling di Sarawak. *Malaysian Journal of Social Sciences and Humanities (MJSSH)*, *8*(1), 1–17. https://doi.org/10.47405/mjssh.v8i1.2038

Yusuf, A. M. (2015). *Asesmen dan evaluasi pendidikan: Pilar penyedia informasi dan kegiatan pengendalian mutu pendidikan* (1st ed.). Prenadamedia Group. https://opac.perpusnas.go.id/DetailOpac.aspx?id=948297

Zhang, Z., & Garcia, L. (2023). Examining dimensionality and validity of the academic integrity survey instrument. *Journal of Education and Development*, *7*(1), 46–51. https://doi.org/10.20849/jed.v7i1.1326