Journal of Culture and Values in Education

https://cultureandvalues.org

E-ISSN: 2590-342X

Volume: 9 Issue: 1 2026

pp. 106-124

Development and Implementation of Smart Measurement and Evaluation (Smart ME): A Trial-by-Trial Monitoring System for Autism Therapy

Arneliza Arneliza[*a], Rudy Sutadi[a], Asmadi Alsa[a] & Kuncono Teguh Yunanto[a]

* Corresponding author:
Email: arneliza.2366290016@upi-yai.ac.id
a. Faculty of Psychology, Persada Indonesia University- UPI Y.A.I., Jakarta, Indonesia.

Article Info
Received: October 08, 2025
Accepted: February 03, 2026
Published: February 16, 2026

10.46303/jcve.2026.5

How to cite
Arneliza, A., Sutadi, R., Alsa, A., Yunanto, K. T. (2026). Development and Implementation of Smart Measurement and Evaluation (Smart ME): A Trial-by-Trial Monitoring System for Autism Therapy. *Journal of Culture and Values in Education*, 9(1), 106-124. https://doi.org/10.46303/jcve.2026.5

ABSTRACT
In Applied Behavior Analysis (ABA) intervention practice, measurement quality plays a crucial role and is a key foundation, as instructional decision-making relies heavily on accurate, consistent, and transparent behavioral data. However, ABA practice still faces challenges such as inconsistent passing criteria, trial-by-trial response recording, documentation procedures, and limited data transparency in autism therapy services. This study aims to develop and implement Smart Measurement and Evaluation (Smart ME) as a standardized measurement and evaluation system that supports detailed response recording, data-driven instructional decision-making, and transparency in Smart ABA therapy services. The study employed a research and development design. Smart ME was developed through discussions with experts and practitioners, then validated using Aiken's V. Limited implementation was conducted with two children with Autism Spectrum Disorder in a Smart ABA therapy session using a two-on-one format to assess the system's operational feasibility and ease of use. The validation results showed an Aiken's V value of 0.92, indicating high content validity. Limited implementation demonstrates that Smart ME supports consistent trial-by-trial recording, clear session scoring, and easy-to-understand data visualization. Smart ME is a content-valid measurement and evaluation system that is operationally feasible for use in autism therapy practice. Smart ME enhances measurement transparency, supports data-driven instructional decisions, and enhances accountability for autism therapy services.

KEYWORDS
Autism Spectrum Disorder; Smart ABA, Smart Measurement and Evaluation; Evidence-based intervention; Two on one.

INTRODUCTION

Autism is an intricate neurobiological disorder that influences communication skills, social interactions, learning, and behavior, as described in contemporary research (Appah et al., 2024; Blaxill et al., 2025; Salari et al., 2022; Sit & Erkan, 2024; Sithole-Tetani, 2024). With the rising prevalence rate of autism globally, the need for proper measurement and evidence-based intervention systems has become an increasing priority for multiple stakeholders, as clearly advocated by (Hume, 2021; Jabery & Arabiat, 2024; Shire et al., 2021). In this context, measurement quality plays a critical role in determining whether observed behavioral changes genuinely reflect intervention effects or are artifacts of flawed or inconsistently applied measurement procedures.

Applied Behavior Analysis (ABA) is widely recognized as an evidence-based intervention for individuals with autism spectrum disorder and is fundamentally grounded in objective, systematic, and direct behavioral measurement (Cooper et al., 2020; Johnston & Pennypacker, 2008). Within ABA practice, instructional decisions are closely tied to ongoing measurement of observable behavior, as progress data are used to evaluate intervention effectiveness and guide subsequent instructional planning. Internal validity can be considered accurate for ABA intervention outcome studies if interobserver agreement (IOA) and procedural fidelity are established (Morris et al., 2024; Petursdottir & Carr, 2018).

Accordingly, IOA and procedural fidelity function to reduce observer bias and ensure that observed behavioral change can be attributed to intervention effects rather than measurement error. Different forms of measurements in ABA, whether continuous or discontinuous, need to be documented, along with their operational definitions (Ledford & Gast, 2018). In addition, ABA emphasizes the importance of clearly operationalized behavioral definitions and the consistent application of measurement procedures to document behavioral change across sessions. However, there are major challenges in ABA practice, in the form of inconsistencies in the application of mastery criteria, recording procedures, data transparency, and measurement documentation at the practice level.

Despite the central role of measurement within ABA, evidence from both empirical and methodological studies indicates that challenges remain in how measurement is implemented in everyday practice. Previous research has documented substantial variability in mastery criteria, with thresholds ranging from approximately 80% accuracy to perfect performance across consecutive sessions (Charlop-Christy et al., 2000; DiSanti et al., 2020; Smith, 2001). Such variation may lead to differences in how quickly instructional targets are advanced or maintained, potentially influencing the stability and generalization of newly acquired skills.

Beyond mastery criteria, studies have also highlighted inconsistencies in recording procedures, documentation of trial-by-trial responses, and reporting of interobserver agreement and procedural fidelity (Bergmann et al., 2023; Essig et al., 2023; Morris et al., 2024). These issues are particularly salient in one-to-one intervention contexts, where therapists are often responsible for both instruction and data recording. Under such conditions, attentional

demands may increase, and opportunities for systematic verification of measurement accuracy may be limited (Koegel et al., 2012; Leaf et al., 2016; Walton & Ingersoll, 2012). Evidence from implementation science in autism intervention highlights that variability and fragmentation in adopting evidence-based practices across service systems further challenge consistent measurement and integration of standardized evaluation procedures (Lee et al., 2025).

At the service level, findings from rehabilitation and allied health research further suggest that the absence of standardized measurement and evaluation systems constrains longitudinal monitoring and complicates meaningful comparison of intervention outcomes across practitioners and contexts (Bourke-Taylor et al., 2018; Kamper, 2019; Rognstad et al., 2023). Taken together, these findings indicate that current measurement challenges in autism intervention are not rooted in a lack of measurement principles, but rather in how those principles are operationalized, documented, and integrated into routine practice.

Overall, the literature indicates that measurement challenges in ABA practice are predominantly procedural and implementive rather than conceptual. While direct behavioral measurement principles are well established, their application in routine autism intervention settings often remains fragmented, inconsistently documented, and insufficiently integrated into systematic evaluation processes. This condition underscores the need for a structured measurement system that not only records behavioral change but also supports consistent interpretation, verification, and instructional decision-making across sessions and service contexts.

In response to this gap, Smart Measurement and Evaluation (Smart ME) was developed as a structured system designed to operationalize and integrate key measurement and evaluation criteria derived from established ABA measurement practices (Cooper et al., 2020; Johnston & Pennypacker, 2008; Ledford et al., 2008). Although these procedures are well established, previous studies indicate that the implementation and documentation of measurement at the practice level particularly trial-by-trial recording, procedural reporting, and transparency of progress data to parents remain variable and have not been systematically integrated into a coherent evaluation framework (MacFarlane et al., 2023). The Measurement and Evaluation (ME) criteria adapted in this study include: direct measurement of observable behavior, clearly operationalized response definitions, consistent trial-by-trial data recording, explicit evaluation rules linking session data to instructional decisions, and transparency mechanisms that enable verification and meaningful access to progress data by supervisors and parents.

Accordingly, this study aims to: (1) describe the structural components and procedural framework underlying the development of the Smart ME system; (2) examine the content validity of Smart ME based on expert and practitioner evaluations; and (3) evaluate the feasibility and usability of Smart ME when implemented in autism therapy sessions. Together, these objectives position the study as a development-focused investigation that addresses

practical measurement needs while strengthening the alignment between behavioral data and instructional decision-making in applied autism intervention.

METHODOLOGY

Research Design

This study employed a research and development (R&D) method. This study employed a developmental research design with a research and development (R&D) orientation, focusing on the early-stage development and operational feasibility of a measurement and evaluation system for autism intervention within the Smart ABA framework. The primary aim of this design was not to test intervention effectiveness, but to design, validate, refine, and examine the practical usability of a measurement system intended for routine application in autism therapy settings. Rather than implementing a full-scale R&D model such as Borg and Gall or ADDIE, this study adopted a limited R&D approach, which is consistent with development research that prioritizes product design, content validation, and feasibility testing prior to large-scale effectiveness evaluation (Richey & Klein, 2007).

This approach is appropriate for studies seeking to bridge the gap between established theoretical principles and their systematic implementation in applied practice. The development procedure consisted of four sequential stages. First, the design stage involved identifying measurement and documentation needs in applied autism therapy contexts and constructing the initial version of the Smart Measurement and Evaluation (Smart ME) system. At this stage, the instrument was designed to support trial-by-trial response recording, session-level documentation, and transparent reporting of intervention progress in alignment with established ABA measurement principles.

Second, the content validation stage focused on evaluating the clarity, relevance, and appropriateness of the Smart ME system. Validation procedures examined the suitability of operational definitions, response recording formats, evaluation rules, and the language used throughout the system. Expert judgment and practitioner feedback were employed to assess whether the instrument content adequately represented measurement requirements in autism therapy practice and could be understood and applied consistently by users.

Third, the revision stage involved refining the Smart ME system based on feedback obtained during content validation. Revisions addressed issues related to the clarity of instructions, consistency of documentation procedures, and the practicality of implementing the system during ongoing therapy sessions. This stage ensured that the system was not only conceptually sound but also feasible for use under typical clinical conditions.

Finally, the implementation and feasibility testing stage involved applying the revised Smart ME system in autism therapy sessions. This stage aimed to evaluate the operational feasibility of the system, including the clarity of trial-by-trial recording procedures, transparency of session-level reporting, consistency of documentation across sessions, and ease of use by therapists during routine intervention practice. The focus of this stage was on usability and

practicality rather than on measuring intervention outcomes. Accordingly, the present study is positioned as a development-focused investigation that examines the feasibility and practical implementation of a measurement and evaluation system. It is not intended as a descriptive quantitative study nor as a full experimental evaluation of intervention effectiveness, but rather as an early-stage development study designed to strengthen the alignment between behavioral measurement practices and instructional decision-making in applied autism intervention contexts.

Participants and Setting

In the limited implementation stage, two children diagnosed with ASD by medical professionals in Indonesia, aged seven years (A) and six years (B), were involved. Participants were selected purposively. The inclusion criteria were: (1) a confirmed clinical diagnosis of ASD, (2) active participation in Smart ABA therapy, (3) ability to engage in structured therapy sessions, and (4) consistent attendance during the implementation period. The exclusion criteria included: (1) sensory, motor, or severe medical conditions that could hinder participation in therapy sessions, and (2) inconsistent attendance that might affect the accuracy of trial-by-trial data recording.

Parents were provided with a detailed explanation of the implementation procedures and gave written informed consent. This limited implementation phase aimed to examine the procedural feasibility, implementability, and observability of the Smart Measurement and Evaluation (Smart ME) system in routine therapy sessions. The study was conducted at a Smart ABA therapy service center in Indonesia, in separate therapy rooms equipped with a Closed-Circuit Television (CCTV) system. During the sessions, parents were able to monitor the entire process from outside the therapy room via CCTV for transparency and verification purposes, as shown in Figure 1.

Figure 1.

Smart ABA therapy room & CCTV monitor for parents

Development & Measures

In this study, development was carried out through group discussions. In these group discussions, three therapists, one supervisor, one program director, and one consultant discussed the development design step by step. The discussion began with designing a systematic, measurable, easy-to-use, and accessible instrument for therapists, parents, and professionals working in this field. After several stages of discussion, the team agreed to name the Smart Measurement and Evaluation, abbreviated to Smart ME.

The Smart ME draft was tested for readability by five therapists who treat children with ASD. This was then followed by assessments by experts and practitioners (three experts in the fields of Education, Psychology, and Language Studies, and two practitioners who treat autism). The expert and practitioner assessments included (1) relevance and appropriateness of the content, (2) ease of use of the content, and (3) appropriateness of language use. This assessment was carried out using a five-point scale (very relevant, relevant, somewhat relevant, not relevant, and very irrelevant). The results of the expert and practitioner assessments were then analyzed using Aiken's V formula (Aiken, 1985).

***Conceptualization* of the Smart ME**

In this study, the term "item" does not refer to questionnaire statements, but to clearly defined operational response units and evaluation rules applied at the trial level. Each response category (e.g., X, –, O, P+, +, A), cycle structure, and scoring rule constitutes a measurement item that was subjected to content validation. Accordingly, item statements in Smart ME are presented as explicit behavioral definitions and procedural rules, rather than attitudinal statements or numerical indicators alone. These measurement items differ from conventional ABA measurement approaches, which primarily rely on binary correct-incorrect scoring. In contrast, Smart ME incorporates multi-category response classification, trial-by-trial recording, and explicit evaluation rules to guide instructional decision-making.

Unlike conventional ABA therapy, which is typically conducted by a single therapist and may present challenges in accurate response recording (Saint-Georges et al., 2020; Wood et al., 2021), Smart ABA is implemented using a two-therapist (two-on-one) model. In this arrangement, the therapist and therapist assistant alternate roles during each session. The therapist assistant supports session preparation by helping the child remain seated and attentive, provides prompts when necessary, and records all child responses on a trial-by-trial basis during instruction.

Recording responses at the trial level supports greater procedural consistency and reduces the risk of observational bias. When instructional delivery and response recording are handled by the same individual, measurement accuracy may be compromised due to increased attentional demands. By separating instructional and recording roles and embedding these roles within a structured measurement system, Smart ABA enhances the reliability and authenticity of behavioral data collected during therapy sessions (Bergmann et al., 2023; Gardner et al., 2013; Morris et al., 2024).

In Smart ABA, on non-verbal programs, when a child is given an instruction, there are five possible responses, each coded with a specific sign: incorrect (sign X), no response (sign –), off-task (sign O), partially correct (sign P+), and correct (sign +), correct response to the third instruction, which was immediately prompted by the therapist assistant, was coded with a P (Anwar et al., 2022).

In the verbal imitation program (words/sentences), responses that approximated the target behavior were coded as A (approximation). In contrast, in the labeling and question-

answering programs, near-correct responses were coded as + (plus). The therapist assistant carefully documented all approximate responses of the child. After completing each therapy session, the therapist assistant calculated the percentage score for all programs within that session. The score was calculated using the following formula:

$$\text{Score} = \frac{\text{Correct Response (CR)}}{\text{Oportunity (OP)}} \text{ x } 100\%$$

CR (Correct Response) refers to the number of correct answers given by the child, and OP (Opportunity) refers to the number of opportunities given to the child to answer. To simplify calculating OP, all answers are counted except those coded P. The scores were then plotted on graphs. The passing criterion was set as achieving a score of ≥ 80% in three consecutive sessions, and a fail system was applied. Each session consisted of three to five cycles, except under specific conditions, such as a consultant's recommendation to perform more than five cycles, which was allowed.

There are three types of cycles in Smart ABA:

Complete cycle: For the first and second instructions, the child provides a response of X, –, or O. On the third instruction, the child receives a prompt from the therapist assistant (examples: XXP, --P, OOP, -XP, O-P, OXP);

Two-instruction cycle: On the first instruction, the child responds with X, –, O, or P+, whereas on the second instruction, the response is + or P+;

One-instruction cycle: The child provides a response of + or P+ on the first instruction. Each program has measurement sheets and graphs, for example, if there are 17 programs given to children, there must be 17 measurement sheets and 17 graph sheets available. An example of Smart ME implementation in Smart ABA therapy is shown in Table 1 below.

Table 1.

Response Recording and Scoring

		TRIAL										**TOTAL**		
Date	**Session**	1	2	3	4	5	6	7	8	9	10	**Op**	**CR**	**Score**
1	1	X	X	P	X	X	P	X	X	P	X	10	0	0%
		X	P	X	X	P								
1	2	X	X	P	X	X	P	-	O	P	X	15	3	20%
		X	P	X	P+	X	X	P	+	+	+			
1	3	-	X	P	+	+	P+	-	+	+	P+	13	8	61%
		+	+	+	+									
1	4	+	+	O	+	+	+	+				7	6	85%
2	1	+	+	+	+	+						5	5	100 %
2	2	X	+	+	X	+	+	+	+	+		9	7	77%
2	3	+	+	+	+							4	4	100%
2	4	+	+	O	+	+	+	+	+			8	7	87%
3	1	+	+	+	+	+						5	5	100%

Fall Passed/Achieved

Table 1 shows that more than five cycles were sometimes conducted. This occurred due to the implementation of Smart EO (Smart Establishing Operation). Smart EO is applied when, within a Smart DTT sequence, the child provides a correct response after previously responding with X, –, or O. For Smart EO, therapists randomly repeat the program 1-3 times, then move on to another program and return to the core program, but Smart EO is beyond the scope of this study. The therapist assistant then plots the total score on a graph at each session. The use of this graph is crucial, as it serves as a guide for therapists and therapist assistants to analyze progress, stagnation, or regression in a program. Smart ME is used in all programs, both daily and maintenance.

If there are obstacles/problems in program implementation that result in the child not meeting the graduation criteria, the therapist identifies the problem and reports it in a session report. This reporting of problems is intended to receive recommendations for solutions/modifications from the supervisor, program director, and consultant. However, if the problem is not resolved, the therapist and therapist assistant prepare a recording of the problematic program, to be discussed in a case presentation every Saturday. This presentation is attended by all therapists, supervisors, program directors, consultants, and the Smart ABA support system. An example of the Smart ME graph can be seen in Table 2.

Table 2.

Graph of scoring calculation

Session	1	2	3	4	1	2	3	4	1	2	3	4
Scoring												
100 %												
80 %												
60 %												
40 %												
20 %												
0 %												
Date	1	1	1	1	2	2	2	2	3	3	3	3

Data Collection Techniques

Focus groups were conducted in numerous cycles, or sessions, throughout the development period. Numerous notes and discussions were taken to refine the development results, with several revisions made before finalizing the design and components to achieve conceptual feasibility of the Smart ME system. After meeting feasibility, Smart ME was implemented on a limited basis with two children with ASD. Measurement sheets and graphs were provided for data recording. A therapist assistant recorded the children's behavior/responses from trial to trial in real-time. Parental observation via CCTV monitors ensured that the children's behaviors matched the recorded data. This implementation ensured the procedures were feasible and the ease of use of Smart ME.

Internal Validity and Control of Threats

Researchers monitored the quality of implementation throughout the study to ensure that data collected using Smart ME were accurate, consistent, and reliable. Drawing on the internal validity framework proposed by Shadish et al. (2002), internal validity in this study was conceptualized to assess implementation credibility and procedural fidelity rather than to establish causal inference. The implementation was conducted in a standardized therapy room, and all participating therapists were trained and competent Smart ABA practitioners.

To reduce recording inconsistencies and potential observer-related threats, the roles of therapist and therapist assistant alternated across sessions, allowing mutual monitoring during instructional delivery and response recording. The use of clearly operationalized response symbols, structured recording formats, and multi-level supervision further supported accurate and consistent data collection throughout the Smart ME implementation.

Icon validity in this study refers to the clarity and appropriateness of the visual symbols and response codes used to represent children's behavioral responses at the trial level. In Smart ME, each icon (e.g., X, –, O, P+, +, A) functions as a measurement item with a clearly defined operational meaning rather than as a decorative or numerical marker. Icon validity was established through expert and practitioner review, focusing on whether each symbol accurately represented the intended response category, was easily distinguishable, and could be applied consistently during real-time trial-by-trial recording. This process ensured shared understanding among users, minimized ambiguity during data recording, and enhanced the reliability of behavioral documentation across therapists and sessions.

Data Analysis

After the development phase was completed, expert and practitioner assessments were conducted using Aikens V to determine the content validity of the Smart ME system. The results of these expert and practitioner assessments were used to determine the suitability, clarity of procedures, ease of use, and feasibility of the measurement system. During implementation, data were analyzed to assess consistency and stability in response recording for each trial, total scores, graphic readability, and ease of use of Smart ME. This analysis evaluated the operational feasibility, procedural implementation, and ease of use of Smart ME in autism therapy practice.

Ethical Considerations

All research procedures were conducted in accordance with the institutional guidelines and regulations of the university and were approved by the Ethics Committee of ….

RESULTS

1. The Initial Draft and Blueprint of Smart ME Measurement

The initial draft of Smart ME was developed on the basis of the measurement principles in ABA and the practice needs identified through group discussions. These discussions involve therapists, supervisors, program directors, and consultants who are experienced in autism

therapy services. The initial development stage focuses on the preparation of a systematic measurement blueprint that can be applied consistently in therapy sessions.

The Smart ME blueprint includes the determination of response categories, trial-by-trial recording procedures, learning cycle structure, scoring rules, and graph visualization mechanisms at the session level. Each trial is designed to integrate the therapist's instruction, the classification of the child's responses, the calculation of scores, and the presentation of graphs. Thus, behavioral data at the trial level can be accurately summarized at the session level and used as the basis for instructional decision-making.

2. Results of Content Validation by Experts and Practitioners

Once the initial draft was developed, Smart ME was subjected to a content validation process involving three experts from the fields of education, psychology, and language studies, as well as two practitioners of autism therapy. In addition, five Smart ABA therapists were engaged to assess the readability and clarity of the system in the context of daily use in the therapy room.

The aspects evaluated included the relevance and suitability of the content, clarity of the response symbols, ease of implementation of the trial-by-trial recording procedure, and accuracy of the language used in the system. The results of expert and practitioner assessments show that Smart ME content meets the criteria of adequate content validity, with an overall Aiken's V score of 0.92. These findings suggest that the Smart ME framework, response categories, and procedural guidelines are sufficiently clear and feasible for application in the context of autism therapy.

3. Item Quality Based on Measurement and Evaluation Criteria

In this study, the term *"item" does* not refer to questionnaire statements but to operational response units and evaluation rules applied at the trial level. Therefore, the quality of Smart ME items was evaluated based on the operational definition's clarity, the response categories' differences, and the consistency of their application in recording the child's behavior during the therapy session. The evaluation results showed that each response category could be clearly distinguished, applied consistently in trial-by-trial logging, and helped minimize ambiguity in behavioral documentation. Thus, Smart ME items meet the ME criteria in terms of clarity, usability, and procedural consistency.

4. Final version of the Smart ME Response Item

Based on the initial draft and content validation results, the final version of Smart ME establishes the following five main response categories for non-verbal programs:

(1) True (+),
(2) partially true (P+),
(3) Silent/no response (–),
(4) Off-task (O), and
(5) false (X).

An additional category of approximate response (A) is used for verbal programs (imitation of words or sentences), which is when the child produces a response that is close to the target.

This measurement structure differs from that of conventional ABA practices, which generally use only two categories of responses (true and false). Smart ME allows for more detailed recording of child behavior without abandoning the trial-by-trial measurement principles that ABA is based on by incorporating graded response categories and verbal approximations.

5. Results of Smart ME Implementation

After meeting the content validity criteria, a Smart ME was implemented in a limited manner in two children with ASD in a Smart ABA therapy session. The implementation involved four certified Smart ABA therapists in a *two-on-one format*. This stage assesses procedural feasibility, trial-by-trial recording consistency, and reporting clarity at the session level, not to test the intervention's clinical effectiveness.

The implementation results showed that the measurement blueprint and Smart ME icon system could be consistently applied during the therapy session. Trial-by-trial response recording is performed according to established operational definitions, and charts at the session level provide a clear visualization of the child's performance over time.

To illustrate, Smart ME is applied to non-verbal programs that target the execution of simple commands, namely: (1) greeting, (2) throwing, (3) *kiss-bye*, (4) clapping, and (5) closing greeting. Detailed examples of note-taking, score calculations, and session graphs for subjects A and B are presented in the Supplemental Material.

6. Smart ME Characteristics Based on Implementation Results

Smart Measurement and Evaluation (Smart ME) was applied consistently to both daily and maintenance programs using the same recording and reporting procedures. The application of structured trial-by-trial documentation, accompanied by graphical visualization at the session level, allows therapists and therapist assistants to monitor the child's development, identify potential obstacles, and document instructional problems in a timely manner during the therapy process.

Based on the limited implementation results, some of the key characteristics of Smart ME that distinguish it from conventional ABA measurement systems can be identified, namely: (1) the use of multi-level response categories that replace binary right-false assessments, (2) the systematic application of response recording at the trial-by-trial level, (3) the separation of instructional roles and data recording through a two-on-one model, (4) the use of session-level graphs as a tool instructional monitoring, and (5) documentation that supports transparency, supervision, and instructional decision-making processes.

Overall, the results show that Smart ME is operationally feasible as a trial-by-trial-based measurement and reporting system in the context of autism therapy. Detailed examples of Smart ME implementation, including trial-by-trial log sheets, scoring tables, and session-level graphs for Subjects A and B, are presented in the Supplementary Materials (Supplementary Figure S1-S3: Implementation of Smart ME for Subject A; Supplementary Figure S4-S8: Implementation of Smart ME for Subject B). This supplement material is provided as an

illustration of Smart ME implementation in therapy sessions and is not intended as a result of an intervention effectiveness test.

DISCUSSION

The results of this development study show that Smart Measurement and Evaluation (Smart ME) meets the criteria for adequate content validity, as shown by Aiken's V value of 0.92. These findings indicate that Smart ME content including measurement blueprints, response categories, record symbols, and procedural rules is relevant, clear, and appropriate for use by experts and practitioners in the context of autism therapy. According to the literature on instrument development, content validity is an important initial stage to ensure the suitability of the system content with the measurement objectives and the behavioral domains to be represented before further testing (Aiken, 1985; DeVellis, 2017).

In line with the objectives of this study, the validity test on Smart ME focuses on the validity of the content and operational feasibility, not on the validity of the construct or statistical reliability. This approach is consistent with the characteristics of early-stage development research, which aims to ensure that the measurement system has a clear conceptual foundation and an understandable procedural structure and can be applied consistently in practice before broader psychometric testing is conducted. Therefore, the results of this study are not intended to conclude the psychometric quality of Smart ME as a whole, but rather as a starting basis for testing the construct's validity and reliability in future research. Thus, the findings of this study need to be understood within the framework of early-stage development research, which emphasizes content suitability and implementation feasibility as prerequisites before further psychometric testing is conducted.

Operationally, Smart ME supports recording the responses of children on a trial-by-trial basis, session scoring, and data presentation through structured graphical visualization for daily programs and maintenance. The limited implementation findings suggest that the Smart ME recording and reporting system can be consistently implemented in Smart ABA therapy sessions, thus facilitating real-time monitoring of the intervention process. This is relevant to findings in the ABA literature that emphasize the importance of clear, sensitive, and standardized measurement systems to support data-driven decision-making (DDM) (Falligant et al., 2021; Fiske et al., 2021). In applied behavior analysis, systematic and continuous measurement is widely recognized as a prerequisite for accountability and instructional quality, particularly in routine clinical services (Cooper et al., 2020; Normand & Fossa, 2021).

In ABA practice, a single therapist often faces operational constraints, such as documentation delays and increased cognitive burden, which can decrease measurement accuracy. The two-on-one model used in Smart ABA allows for a separation of roles between instruction delivery and data logging, thus supporting more timely and consistent trial-by-trial recording. Previous studies have suggested that role sharing can improve data sensitivity and consistency of behavioral documentation between sessions (Beaulieu et al., 2022; Kelly &

Tincani, 2022). Previous literature has also noted that separating instructional and data-recording roles can reduce practitioner cognitive load and enhance documentation accuracy in ABA settings (Brodhead & Higbee, 2022; Lotfizadeh et al., 2025).

Smart ME also emphasizes the use of clear operational definitions, explicit scoring rules, and data visualization at the session level. This approach is in line with ABA principles that place continuous measurement as the foundation of evaluation of the intervention process, rather than just subjective reporting by practitioners (Cox et al., 2022; Fisher et al., 2023; Leaf et al., 2022). Prior ABA studies have applied varying mastery criteria depending on intervention goals and skill complexity. Early Lovaas-based programs reported highly stringent accuracy standards (Smith, 2001), whereas more recent studies commonly adopted mastery thresholds around 80% across consecutive sessions (DiSanti et al., 2020). With an explicitly defined graduation rule (≥80% for three consecutive sessions), Smart ME also contributes to data interpretation consistency amid variations in mastery criteria reported in previous ABA practice (Grow & LeBlanc, 2013; Ledford & Gast, 2018; Pitts & Hoerger, 2021).

From a theoretical perspective, the findings of this study strengthen the ABA framework, which considers measurement as the main infrastructure in data-driven instructional decision-making (Johnston & Pennypacker). Smart ME clarifies the relationship between the operational definition of behavior, trial-by-trial recording, and therapeutic process evaluation as a single measurement system. This study confirms that the quality of the intervention is determined not only by the teaching procedure but also by a measurement system capable of accurately, consistently, and transparently documenting the therapeutic process.

Practically, the results show that Smart ME can be used as a structured and operationally robust measurement and evaluation system in Smart ABA implementation, providing a strong foundation for systematic documentation and data-driven instructional decision-making. The involvement of parents as observers through the CCTV system also supports the transparency and accountability of services without interfering with the instructional process. Recent discussions in ABA services have emphasized transparency and parent access to intervention processes as an important component of service accountability (Ferretti et al., 2025; Liao et al., 2022). However, testing of construct validity, inter-assessor reliability, and cross-context measurement stability is still needed and recommended as a further research agenda to strengthen the Smart ME's psychometric basis.

CONCLUSION

This study concludes that Smart ME meets the criteria of adequate and operationally feasible content validity as an early-stage ME system in the implementation of Smart ABA therapy. The validity of the obtained content showed the compatibility between the measurement blueprint, response categories, recording symbols, and procedural rules with the purpose of behavioral measurement in the context of autism therapy. As early-stage development research, this study's validation focused on the validity of the content and the feasibility of implementation,

not on the validity of the construct or statistical reliability. Therefore, the results of this study serve as a starting basis for advanced psychometric testing in future research. Smart ME supports trial-by-trial response logging, session-level scoring, data visualization, and structured reporting. Limited implementation shows that this system can be applied consistently and supports the systematic monitoring of the therapeutic process. Further research is recommended to test the Smart ME's construct validity and reliability in a broader context and sample.

Limitations and recommendations

This study has limitations because reliability and procedural fidelity have not been tested quantitatively, and the implementation of Smart ME is still limited to specific contexts and samples. Therefore, further research is recommended to test inter-recorder reliability, measure fidelity independently, and apply Smart ME to a broader sample and service setting to strengthen empirical evidence and support the system's generalizability in Smart ABA practice.

REFERENCES

Aiken, L. R. (1985). Three Coefficients For Analyzing The Reliability And Validity Of Ratings. *Educational And Psychological Measurement*, *45*(1), 131–142. https://doi.org/10.1177/0013164485451012

Anwar, A., Sutadi, R., & Miranda, C. (2022). Development of Discrete Trial Training (DTT) Procedure in Smart Applied Behavior Analysis (Smart ABA) for Autism. *Journal of Psychology and Behavior Studies*, *2*(1), 63–44. https://doi.org/10.32996/jpbs.2022.2.1.7

Appah, J., Senoo-Dogbey, V. E., Armah, D., Wuaku, D. A., & Ohene, L. A. (2024). A qualitative enquiry into the challenging roles of caregivers caring for children with Autism Spectrum Disorders in Ghana. *Journal of Pediatric Nursing*, *76*, 23–29. https://doi.org/10.1016/j.pedn.2024.01.029

Beaulieu, L., Addington, J., & Almeida, D. (2022). Data-based decision making and program monitoring practices in applied behavior analysis. *Behavior Analysis in Practice*.

Bergmann, S., Long, B. P., St. Peter, C. C., Brand, D., Strum, M. D., Han, J. B., & Wallace, M. D. (2023). A detailed examination of reporting procedural fidelity in the Journal of Applied Behavior Analysis. *Journal of Applied Behavior Analysis*, *56*(4), 708–719. https://doi.org/10.1002/jaba.1015

Bergmann, S., Niland, H., Gavidia, V. L., Strum, M. D., & Harman, M. J. (2023). Comparing Multiple Methods to Measure Procedural Fidelity of Discrete-trial Instruction. *Education and Treatment of Children*, *46*(3), 201–220. https://doi.org/10.1007/s43494-023-00094-w

Blaxill, M., Rogers, T., & Nevison, C. (2025). Autism Tsunami : the Impact of Rising Prevalence on the Societal Cost of Autism in the United States. *Journal of Autism and Developmental Disorders*, *52*(6), 2627–2643. https://doi.org/10.1007/s10803-021-05120-7

Bourke-Taylor, H. M., Brown, T., & Cordier, R. (2018). Special Issue: Innovations in occupational therapy measurement. *Australian Occupational Therapy Journal*, *65*(5), 343–345. https://doi.org/10.1111/1440-1630.12521

Brodhead, M. T., & Higbee, T. S. (2022). Practical Considerations for Measurement and Documentation in ABA. *Behavior Analysis in Practice*, *15*, 34–46.

Charlop-Christy, M. H., Le, L., & Freeman, K. A. (2000). A Comparison of Video Modeling with In Vivo Modeling for Teaching Children with Autism. *Journal of Autism and Developmental Disorders*, *30*(6), 537–552. https://doi.org/10.1023/A:1005635326276

Cooper, J. O., Heron, T. E., & Heward, W. L. (2020). *Applied behavior analysis*. Pearson UK.

Cox, D. J., & al., et. (2022). Visual Analysis and Graphing Practices in Behavior-Analytic Research. *Journal of Applied Behavior Analysis*, *55*, 789–803.

DeVellis, R. F. (2017). *Scale Development: Theory and Applications* (4th ed.). SAGE Publications, Inc. https://us.sagepub.com/en-us/nam/scale-development/book243810

DiSanti, B. M., Eikeseth, S., Eldevik, S., Conrad, J. M., & Cotter-Fisher, K. L. (2020). Comparing structured mix and random rotation procedures to teach receptive labeling to children with autism. *Behavioral Interventions*, *35*(1), 38–56. https://doi.org/10.1002/bin.1694

Disanti, J., & al., et. (2020). Mastery criteria and instructional decision making in applied behavior analysis. *Behavior Analysis in Practice*. https://doi.org/10.1007/s40617-019-00398-9

Essig, L., Rotta, K., & Poling, A. (2023). Interobserver agreement and procedural fidelity: An odd asymmetry. *Journal of Applied Behavior Analysis*, *56*(1), 78–85. https://doi.org/10.1002/jaba.961

Falligant, J. M., Higbee, T. S., & Kollins, S. H. (2021). Program monitoring and data-based decision making in applied behavior analysis. *Behavior Analysis in Practice*.

Ferretti, L. A., Uhl, A., Zawacki, J., & McCallion, P. (2025). Understanding Barriers and Facilitators of Parent/Caregiver Involvement in Home-Based Applied Behavioral Analysis Programming for Their Autistic Child. *Children*, *12*(7), 850. https://doi.org/10.3390/children12070850

Fisher, W. W., & al., et. (2023). Improving Reproducibility and Accountability in Applied Behavior Analysis. *Journal of Applied Behavior Analysis*, *56*, 1–15.

Fiske, K. E., Delmolino, L., & Sloman, K. N. (2021). Clinical documentation and program monitoring practices in applied behavior analysis services. *Behavior Analysis in Practice*.

Gardner, L. M., Murphy, L., Campbell, J. M., Tylavsky, F., Palmer, F. B., & Graff, J. C. (2013). Screening accuracy for risk of autism spectrum disorder using the Brief Infant-Toddler Social and Emotional Assessment (BITSEA). *Research in Autism Spectrum Disorders*, *7*(5), 591–600. https://doi.org/10.1016/j.rasd.2013.01.004

Grow, L. L., & LeBlanc, L. A. (2013). Teaching Receptive Language Skills: Recommendations for Mastery Criteria. *Behavior Analysis in Practice*, *6*, 56–65.

Hume, K. (2021). Evidence-Based Practices for Children, Youth, and Young Adults with Autism:

Third Generation Review. *Journal of Autism and Developmental Disorders*, *51*(11), 4013–4032. https://doi.org/10.1007/s10803-020-04844-2

Jabery, M.A.AL & Arabiat, D.H. (2024). Nurturing School Well-being: Insightful Perspectives on Key School Features for Students with Autism, *Journal of Social Studies Education Research, 15*(3), 88-117. https://jsser.org/index.php/jsser/article/view/5725/681

Johnston, J. M., & Pennypacker, H. S. (2008). *Strategies and Tactics of Behavioral Research (3rd ed.)*. Routledge.

Kamper, S. J. (2019). Fundamentals of Measurement: Linking Evidence to Practice. *Journal of Orthopaedic & Sports Physical Therapy*, *49*(2), 114–115. https://doi.org/10.2519/jospt.2019.0701

Kelly, M. P., & Tincani, M. (2022). Reducing Practitioner Burden While Maintaining Data Quality in ABA Practice. *Behavior Analysis in Practice*, *15*, 567–577.

Koegel, R. L., Matos-Freden, R., Lang, R., & Koegel, L. K. (2012). Intervention practices for children with autism: Issues in implementation and fidelity. *Journal of Positive Behavior Interventions*, *14*(3), 166–178. https://doi.org/10.1177/1098300712437044

Leaf, J. B., & al., et. (2022). Ensuring Transparency and Accountability in ABA Service Delivery. *Behavior Analysis in Practice*, *15*, 905–918.

Leaf, J. B., Aljohani, W. A., & Milne, C. M. (2016). Procedural fidelity in applied behavior analysis. *Behavior Analysis in Practice*. https://doi.org/10.1007/s40617-016-0125-1

Ledford, J. R., & Gast, D. L. (2018). *Single Case Research Methodology*. Routledge.

Ledford, J. R., Gast, D. L., Luscre, D., & Ayres, K. M. (2008). Observational and Incidental Learning by Children with Autism during Small Group Instruction. *Journal of Autism and Developmental Disorders*, *38*(1), 86–103. https://doi.org/10.1007/s10803-007-0363-7

Lee, J. D., Terol, A. K., Tschida, J. E., Pomales-Ramos, A., McEathron, S., Wallisch, A., & A Boyd, B. (2025). Examining the use of implementation science in autism intervention research: A scoping review. *Autism*, *29*(12), 2911–2926. https://doi.org/10.1177/13623613251351344

Liao, Y., Dillenburger, K., & Hu, X. (2022). Behavior analytic interventions for children with autism: Policy and practice in the United Kingdom and China. *Autism*, *26*(1), 101–120. https://doi.org/10.1177/13623613211020976

Lotfizadeh, A. D., McCoy, E., & Rico, C. (2025). A Comparison of Electronic and Pen-And-Paper Recording in ABA Sessions. *Behavioral Interventions*, *40*(2), e70013. https://doi.org/10.1002/bin.70013

MacFarlane, H., Salem, A. C., Bedrick, S., Dolata, J. K., Wiedrick, J., Lawley, G. O., Finestack, L. H., Kover, S. T., Thurman, A. J., Abbeduto, L., & Fombonne, E. (2023). Consistency and reliability of automated language measures across expressive language samples in autism. *Autism Research*, *16*(4), 802–816. https://doi.org/10.1002/aur.2897

Morris, C., Jones, S. H., & Oliveira, J. P. (2024). A Practitioner's Guide to Measuring Procedural Fidelity. *Behavior Analysis in Practice*, *17*(2), 643–655. https://doi.org/10.1007/s40617-

024-00910-8

Morris, S. B., DeShon, R. P., & Kelley, K. (2024). Fidelity measurement in behavioral intervention research. *Behavior Analysis in Practice*. https://doi.org/10.1007/s40617-023-00820-9

Normand, M. P., & Fossa, A. A. (2021). Procedural Integrity and Measurement Accuracy in Applied Behavior Analysis. *Behavior Analysis in Practice*, *14*, 1016–1026.

Petursdottir, A. I., & Carr, J. E. (2018). Interobserver agreement: Its importance and how to improve it. *Behavior Analysis in Practice*, *11*(1), 13–19. https://doi.org/10.1007/s40617-018-0218-4

Pitts, L., & Hoerger, M. (2021). Mastery criteria and skill maintenance in autism intervention. *Behavior Analysis in Practice*. https://doi.org/10.1007/s40617-020-00479-5

Richey, R. C., & Klein, J. D. (2007). *Design and Development Research: Methods, Strategies, and Issues*. Lawrence Erlbaum Associates.

Rognstad, K., Wentzel-Larsen, T., Neumer, S.-P., & Kjøbli, J. (2023). A Systematic Review and Meta-Analysis of Measurement Feedback Systems in Treatment for Common Mental Health Disorders. *Administration and Policy in Mental Health and Mental Health Services Research*, *50*(2), 269–282. https://doi.org/10.1007/s10488-022-01236-9

Saint-Georges, C., Pagnier, M., Ghattassi, Z., Hubert-Barthelemy, A., Tanet, A., Clément, M. N., Soumille, F., Crespin, G. C., Pellerin, H., & Cohen, D. (2020). A developmental and sequenced one-to-one educational intervention (DS1-EI) for autism spectrum disorder and intellectual disability: A three-year randomized, single-blind controlled trial. *EClinicalMedicine*, *26*, 1–16. https://doi.org/10.1016/j.eclinm.2020.100537

Salari, N., Rasoulpoor, S., Rasoulpoor, S., Shohaimi, S., Jafarpour, S., Abdoli, N., Khaledi-Paveh, B., & Mohammadi, M. (2022). The global prevalence of autism spectrum disorder: a comprehensive systematic review and meta-analysis. *Italian Journal of Pediatrics*, *48*(1). https://doi.org/10.1186/s13052-022-01310-w

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and Quasi-experimental Designs for Generalized Causal Inference*. Houghton Mifflin.

Shire, Stephanie Y, Shih, Wendy, Barriault, Terri, & Kasari, Connie. (2021). Exploring coaching and follow-up supports in community-implemented caregiver-mediated JASPER intervention. *Autism*, *26*(3), 654–665. https://doi.org/10.1177/13623613211066132

Sit, S., & Erkan , N. S. . (2024). The Relationship between Social Support and Stress and Burnout Levels of Mothers of Children Diagnosed with Autism Spectrum Disorder. *Theory and Practice in Child Development, 4*(1), 76–95. https://doi.org/10.46303.tpicd.2024.5

Smith, T. (2001). Discrete Trial Training in the Treatment of Autism. *Focus on Autism and Other Developmental Disabilities*, *16*(2), 86–92. https://doi.org/10.1177/108835760101600204

Sithole-Tetani, N. (2024). Understanding the Difference Between Autism and the Autistic

Spectrum: A Focus on the South African Context. *Research in Social Sciences and Technology, 9*(3), 327-339. https://doi.org/10.46303/ressat.2024.62

Walton, K. M., & Ingersoll, B. R. (2012). Evaluation of a Sibling-Mediated Imitation Intervention for Young Children With Autism. *Journal of Positive Behavior Interventions*, *14*(4), 241–253. https://doi.org/10.1177/1098300712437044

Wood, J. J., Sze Wood, K., Chuen Cho, A., Rosenau, K. A., Cornejo Guevara, M., Galán, C., Bazzano, A., Zeldin, A. S., & Hellemann, G. (2021). Modular cognitive behavioral therapy for autism-related symptoms in children: A randomized controlled trial. *Journal of Consulting and Clinical Psychology*, *89*(2), 110. https://doi.org/10.1037/ccp0000621

Supplementary Materials: Trial-by-Trial Implementation of Smart Measurement and Evaluation (Smart ME) for Participant A

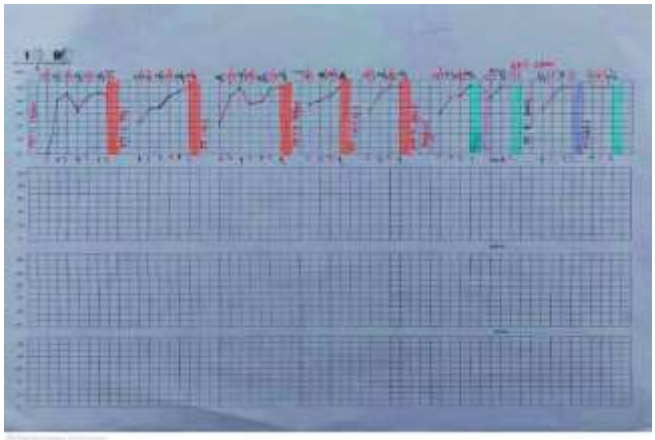
Figure S1 & S2.

Recording and scoring implementation results using the Smart ME system

Recording and scoring implementation results using the Smart ME system. These figures illustrate trial-by-trial response recording and session-level scoring procedures applied during routine Smart ABA therapy sessions. The visualizations demonstrate how responses are documented, aggregated, and displayed according to predefined scoring rules. This supplementary material is provided to clarify the operational use of Smart ME in practice and is not intended to represent intervention effectiveness.

Figure S3.

Trial-by-trial performance graph for Participant A generated by the Smart ME system



The graph presents session-based performance data derived from trial-by-trial recording during

routine Smart ABA therapy sessions. It illustrates how individual responses are plotted and summarized across sessions in accordance with the Smart ME scoring framework. This figure is included to demonstrate the use of graphical visualization for monitoring and documentation purposes and does not represent an evaluation of intervention effectiveness.

Supplementary Materials: Trial-by-Trial Implementation of Smart Measurement and Evaluation (Smart ME) for Participant B

Figure S4 & S7.

Recording and scoring implementation results using the Smart ME system

Figure 8.

Trial-by-trial performance graph for Participant A generated by the Smart ME system