Social Sciences & Humanities Open 13 (2026) 102641

Contents lists available at ScienceDirect

Social Sciences & Humanities Open

journal homepage: www.sciencedirect.com/journal/social-sciences-and-humanities-open

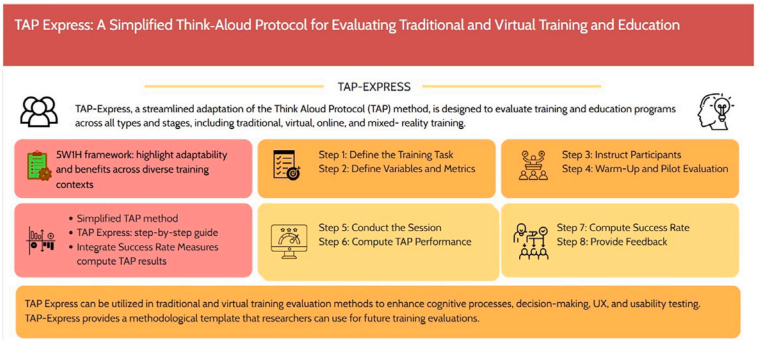
Regular article

TAP express: A simplified think-aloud protocol for evaluating traditional and virtual training and education

Abdallah Al-Hamad[a,*], Attila Gilányi[b]

[a] Doctoral School of Informatics, University of Debrecen, Egyetem tér 1, 4032 Debrecen, Hungary
[b] Faculty of Informatics, University of Debrecen, Egyetem tér 1, 4032 Debrecen, Hungary

GRAPHICAL ABSTRACT



TAP Express: A Simplified Think-Aloud Protocol for Evaluating Traditional and Virtual Training and Education

TAP-EXPRESS

TAP-Express, a streamlined adaptation of the Think Aloud Protocol (TAP) method, is designed to evaluate training and education programs across all types and stages, including traditional, virtual, online, and mixed- reality training.

5W1H framework: highlight adaptability and benefits across diverse training contexts

Step 1: Define the Training Task
Step 2: Define Variables and Metrics

Step 3: Instruct Participants
Step 4: Warm-Up and Pilot Evaluation

- Simplified TAP method
- TAP Express: step-by-step guide
- Integrate Success Rate Measures compute TAP results

Step 5: Conduct the Session
Step 6: Compute TAP Performance

Step 7: Compute Success Rate
Step 8: Provide Feedback

TAP Express can be utilized in traditional and virtual training evaluation methods to enhance cognitive processes, decision-making, UX, and usability testing.
TAP-Express provides a methodological template that researchers can use for future training evaluations.

ARTICLE INFO

Handling Editor: Daria Khlevniuk

Keywords:
Education
Training evaluation
Think aloud protocol
Cognitive skills
Success rate measurement
User experience

ABSTRACT

Traditional Think-Aloud Protocol (TAP) faces limitations in resource-constrained training environments due to its time-intensive setup and complex analysis. To address this, this study introduces TAP-Express, a streamlined adaptation of the TAP method, designed to evaluate training and education programs across non-technological settings (e.g., traditional classrooms, workshops, and field exercises) and technological modalities (e.g., virtual reality, online platforms, and augmented reality simulations). TAP-Express simplifies the classic TAP method by focusing on key cognitive and intrapersonal skills during training tasks and managing the required verbalization. The methods used to ensure robust evaluation across training contexts and objectives include providing a replicable step-by-step framework, integrating success rate measurement, and supporting small-scale studies. The methodology employs an eight-step process with standardized tools (e.g., task definition tables, pilot study templates). TAP-Express enhances accessibility for diverse researchers and offers a practical, inclusive evaluation tool as a methodological template for education and healthcare training, thereby contributing to the social sciences by combining qualitative and quantitative insights.

---

* Corresponding author.
Email addresses: abdallah001@mailbox.unideb.hu (A. Al-Hamad), gilanyi@inf.unideb.hu (A. Gilányi).

https://doi.org/10.1016/j.ssaho.2026.102641
Received 8 April 2025; Received in revised form 18 February 2026; Accepted 27 February 2026

1. Introduction

Training and education evaluation are critical to organizational development, ensuring that training programs effectively enhance employee skills and performance. Non-technological (traditional) methods, such as surveys and post-training assessments, often fail to capture real-time cognitive processes of trainees. The Think-Aloud Protocol (TAP) (Ericsson & Simon, 1993) addresses this gap by providing insights into trainees' thought processes during task execution. However, traditional TAP's limitations, including time-intensive setup, extensive participant training, and complex data analysis, restrict its use in resource-constrained environments such as online education or virtual reality (VR) simulations. TAP-Express comes with innovative methodologies that streamlines TAP to focus on critical decision points, enhancing efficiency and accessibility for evaluating both non-technological and technological training contexts.

For robust evaluation, **TAP-Express** focuses on the critical decision points in training tasks through streamlining the TAP method and adding Success Rate Measurement as an observational score. This streamlined protocol addresses TAP's practical limitations, making it suitable for rapid training contexts and interdisciplinary applications, in education, healthcare, and utilizing new technologies (Radhakrishnan et al., 2021). TAP-Express meets the demand for robust, accessible methodologies in social sciences and humanities research. TAP-Express supports pilot testing, small sample sizes, and mixed-methods evaluation, ensuring comprehensive assessment in resource-limited settings.

This study provides a step-by-step guide for implementing TAP-Express and a methodology for evaluating training programs.

This paper aims to:

1. TAP-Express introduces TAP applications comparative analysis in non-technological (traditional) and technological (e.g., virtual, online) training, using the 5W1H framework (Lewrick et al., 2020) to demonstrate adaptability across social science domains. This objective focuses on evaluating decision-making in contexts such as classroom training or VR simulations.
2. Provide a detailed step-by-step guide for TAP-Express implementation, which indicates the TAP-Express practical applicability in real-world training environments with limited time and resources.
3. Integrate the observational method: utilizing Success Rate Measurement to enrich the outcomes of the traditional TAP measurements, which enrich the evaluation framework of the qualitative cognitive insights with quantitative performance metrics.
4. Support small-scale studies and pilot testing, enabling researchers to refine training protocols efficiently before large-scale implementation.
5. This research highlights TAP-Express's role in facilitating pilot testing that allows researchers to test and refine training protocols efficiently before large-scale implementation, supporting the need for a clear methodological justification.
6. Enhance accessibility for diverse researchers, including early-career scholars and underrepresented groups, meeting social sciences and humanities inclusive research goals.

TAP-Express objectives, summarized in Fig. 1, highlight TAP-Express's strength and traditional TAP's limitations, and support practical, inclusive, and comprehensive training evaluation.

1.1. Research questions

To guide the study and address the manuscript's objectives, the following research questions are posed: These questions ensure a focused evaluation of TAP-Express's effectiveness, applicability, and reproducibility.

1. How effective is TAP-Express in capturing cognitive processes during training tasks compared to traditional TAP?
2. Is TAP-Express applicable to both non-technological (traditional) and technological (virtual) training contexts?
3. Can TAP-Express provide a reproducible framework for social science applications, such as education and healthcare?

This research is structured to systematically present TAP-Express and its evaluation as shown in Fig. 2, beginning with an Introduction that outlines the need for improved training evaluation, highlights limitations of traditional Think-Aloud Protocol, and introduces TAP-Express's innovations to set the research context and objectives, followed by a detailed description of the eight-step TAP-Express protocol, including task definition, data collection, and performance metrics, to provide a clear, replicable methodology for researchers, then a pilot study design demonstrating TAP-Express's application through a virtual reality-based surgical simulation to offer a practical, interdisciplinary example, and concluding with a Discussion that synthesizes findings, addresses research questions, and evaluates implications for social sciences to assess the protocol's impact and limitations. This organization ensures a logical flow from problem identification to practical application and evaluation.

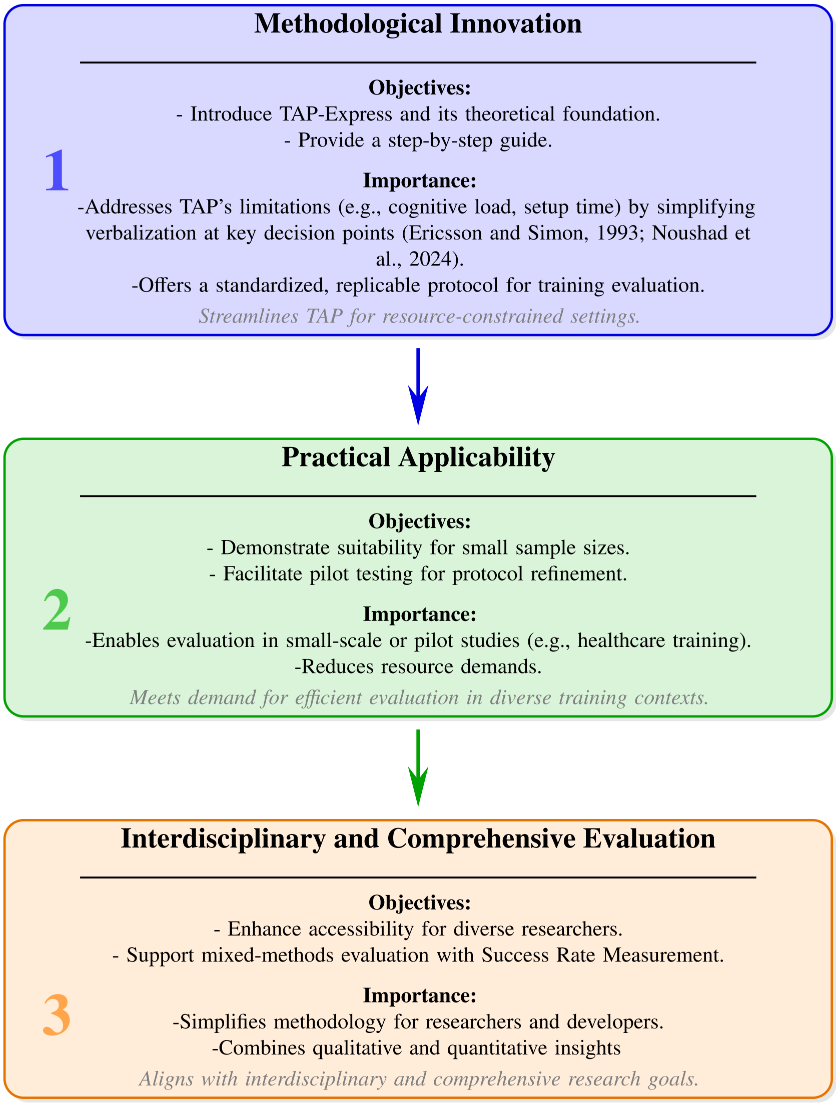
2. Literature review

TAP, introduced by Ericsson and Simon (1980), collects verbalized thoughts during tasks and is widely used in usability testing, education, cognitive psychology, and training evaluation. TAP captures cognitive processes but faces limitations like cognitive load from continuous verbalization (Ericsson & Simon, 1993), recall bias in retrospective methods (Van Den Haak et al., 2003), and inconsistent implementation (Noushad et al., 2024). TAP-Express streamlines the protocol, keeping its key benefits. TAP-Express is utilized to evaluate various social sciences training types, including education (Ciccarelli, 2024), healthcare (Zhao et al., 2024), and psychology (Ericsson & Simon, 1980), ensuring accessible tools for diverse training contexts. This section identifies TAP applications research gaps in the evaluation of both non-technological and technological training.

2.1. Think aloud protocol for training and education evaluation

The way TAP evaluates the effectiveness of training and education was examined in the previous studies. Whereas Eccles explores the application of TAP in various settings (Eccles & Arsal, 2017), additionally, Zhang explores various applications that include cognitive validation, analysis of the writing process, hypothesis generation, medical education, and e-learning environments (Zhang & Zhang, 2019). Furthermore, in cognitive validation, TAP is utilized to validate self-reported questionnaires, ensuring participants interpret questions as intended by researchers. For example, Krishna's research, which explores the TAP function to validate periodontitis assessments, improves the reliability of educational and clinical evaluations (Krishna-Naik et al., 2024).

Additionally, TAP evaluates the process of educational programs, for example, Ciccarelli investigates students' writing to address bureaucratic language tendencies, fostering democratic language education and improved writing outcomes (Ciccarelli, 2024). Moreover, in clinical settings, TAP reveals cognitive events during data analysis, highlighting hypothesis generation differences between experienced and novice researchers (Jing et al., 2023). In healthcare education, think-aloud online sessions significantly improved clinical skills for the students' grasp, with 91% of participants finding the method effective (Menon et al., 2022). Furthermore, integrating TAP with case-based learning enhances critical thinking and improves case analysis scores among new nurses, fostering deeper cognitive engagement (Zhao et al., 2024). Similarly, TAP streamlines user feedback on persuasive learning tools in e-learning, improving educational tech and student decision-making (Abd Rahman et al., 2023). It acts as a flexible tool, strengthening training methods and evaluation across varied educational contexts.



Methodological Innovation

Objectives:
- Introduce TAP-Express and its theoretical foundation.
- Provide a step-by-step guide.

Importance:
-Addresses TAP's limitations (e.g., cognitive load, setup time) by simplifying verbalization at key decision points (Ericsson and Simon, 1993; Noushad et al., 2024).
-Offers a standardized, replicable protocol for training evaluation.

Streamlines TAP for resource-constrained settings.

Practical Applicability

Objectives:
- Demonstrate suitability for small sample sizes.
- Facilitate pilot testing for protocol refinement.

Importance:
-Enables evaluation in small-scale or pilot studies (e.g., healthcare training).
-Reduces resource demands.

Meets demand for efficient evaluation in diverse training contexts.

Interdisciplinary and Comprehensive Evaluation

Objectives:
- Enhance accessibility for diverse researchers.
- Support mixed-methods evaluation with Success Rate Measurement.

Importance:
-Simplifies methodology for researchers and developers.
-Combines qualitative and quantitative insights

Aligns with interdisciplinary and comprehensive research goals.

Fig. 1. TAP-express objectives and importance.

TAP is traditionally rooted in cognitive psychology and positivist philosophies, which assume that verbalizations provide objective access to cognitive processes as information-processing mechanisms (Ericsson & Simon, 1993). However, TAP can also be adapted to alternative paradigms, such as interpretivism, where it serves to uncover subjective meanings and lived experiences during tasks, aligning with qualitative inquiries into participants' interpretations. Pragmatism further expands TAP's scope by supporting mixed-methods applications, allowing researchers to combine verbal data with observational metrics for practical, context-specific insights in training evaluation. This philosophical flexibility broadens TAP's utility beyond quantification to encompass deeper, contextual understanding in diverse fields such as education and healthcare.

While TAP has traditionally been employed in quantitative frameworks to measure cognitive processes through coding and counting verbalizations, it also lends itself to qualitative analyses. For instance, researchers can apply thematic analysis to identify recurring themes in participants' verbal reports, revealing underlying cognitive strategies and mental models. Interpretive phenomenological analysis (IPA) can further explore the lived experiences embedded in these verbalizations, providing rich, contextual insights into how individuals perceive and navigate tasks. This qualitative lens complements the positivist underpinnings of TAP by emphasizing subjectivity and depth, allowing for a more holistic understanding of cognition in training and education contexts.

TAP has also evolved significantly in the sport domain, extending beyond traditional cognitive validation to support metacognition, reflective practice, and self-regulation in coaching and athlete performance. Stodter and Whitehead (2024) critically examine think-aloud and stimulated recall methods in sport coaching, comparing their practicalities, paradigmatic considerations, and value for qualitative insights into coaches' in-action cognitions and behavioral aspects. Similarly, Birch et al. (2022) offer practical guidelines for think-aloud as a facilitator of self-regulation in golfers, illustrating how verbalizing thoughts during performance enhances awareness, adaptive strategies, and reflective processes. These advancements highlight TAP's adaptability in high-pressure, performance-oriented contexts, providing transferable insights for evaluating training protocols in education, healthcare, and virtual simulations through real-time cognitive capture.

2.2. Think aloud protocol for online and virtual training

Utilization of TAP in online, VR, literature, usability, and cognitive reflection assessments, while virtual think-aloud protocols support

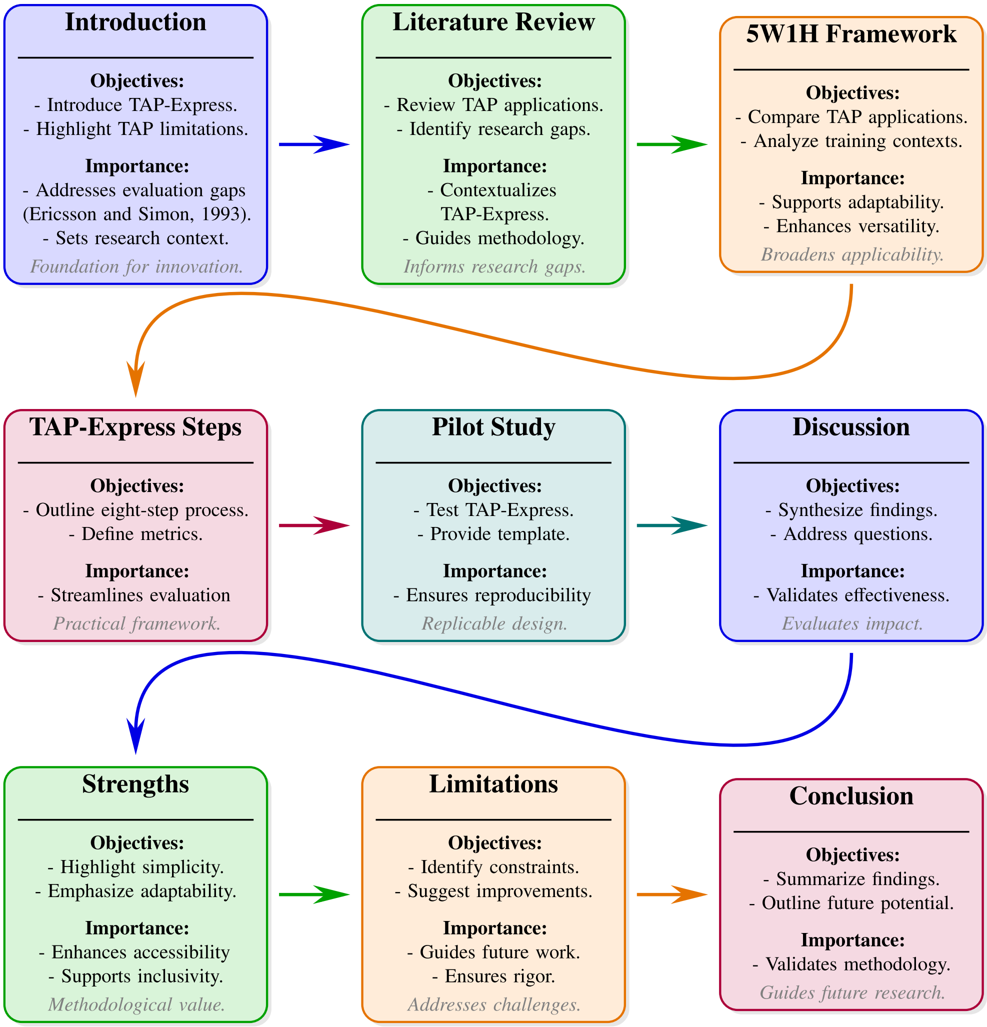


Fig. 2. Research framework.

qualitative research with flexible scheduling, enhanced notetaking, and efficient virtual interviews, particularly in computer science education (Imbulpitiya et al., 2023).

TAP evaluates virtual prototypes in immersive VR environments, with Zhang and Simeone (2022) identifying over 60% of usability issues common to physical and virtual prototypes, demonstrating its effectiveness in virtual training contexts. Other research tackles TAP in VR safety training, where TAP captures trainees' cognitive processes, providing insights into task performance, user experience limitations, and learning opportunities (Yang et al., 2024; Zhang & Simeone, 2022). It integrates into VR settings for healthcare, allowing professionals to refine technical and non-technical skills without ongoing expert supervision (Negrão et al., 2023). Additionally, TAP reveals decision-making processes in cognitive reflection tests, confirming its utility in assessing reflective thinking (Byrd et al., 2023). The synthesis of TAP in virtual training contexts underscores its ability to enhance learning outcomes and cognitive engagement across disciplines (Marijuan, 2024).

2.3. Research gaps

This research identifies the following gaps:

- **Limited Standardization in TAP Utilization**: TAP lacks systematic frameworks for consistent implementation in traditional and virtual training, reducing study comparability. The guidelines of Noushad aid TAP adaptation (Noushad et al., 2024), yet variability remains. This study proposes a structured, adaptable TAP approach.
- **Insufficient Integration with Behavioral Observations:** Most TAP evaluations prioritize verbalized thoughts, neglecting crucial non-verbal cues like behavioral reactions, error frequencies, and hesitation patterns in training assessments.
- **Narrow Skill Assessment Criteria:** Current TAP-based studies evaluate a few skills. A systematic approach integrating cognitive, behavioral, and performance-based criteria is required for comprehensive training assessment.
- **Absence of a Standardized Success Measurement Framework:** No standard template measures success in traditional and VR training. This study presents a framework with performance metrics such as task completion, error correction, and efficiency.

Addressing research gaps enhances TAP's reliability and utility for evaluating traditional and virtual training, enabling comprehensive assessments of training effectiveness across diverse domains.

3. 5W1H framework: a comparative analysis of TAP in non-technological and technological training

TAP's applications in traditional and virtual training evaluation are analyzed using the 5W1H framework (What, Who, Where, Why, When, How) (Lewrick et al., 2020). This structured approach, detailed in Table 1, supports systematic assessment across diverse training methods.

Table 1
5W1H framework for TAP applications in non-technological and technological training.

Aspect	Non-technological Training (Traditional)	Technological Training (Virtual)
What	Cognitive and Intrapersonal skills (Noushad et al., 2024) Decision-making, and problem-solving (Zhang & Zhang, 2019) Education (Ciccarelli, 2024) Skills enhancement (Jing et al., 2023) Psychology and Human Factors (Jääskeläinen, 2012)	In addition to the goals of traditional training: HCI (Marco-Ruiz et al., 2017) Usability Testing (Hertzum, 2024) Training and Simulation (Afifi et al., 2022) Design Research (Negrão et al., 2023)
Who	Students, Trainees, Instructors, Professionals, Researchers	Users, Designers, Developers, Researchers
Where	Classrooms, Workshops, Laboratories	Websites, Apps, VR/AR Systems, Wearable Devices, Metaverse
Why	Enhance Training and Education Improve Cognitive and Intrapersonal Skills Cognitive Load Management Forensic Psychology Behavioral analysis Performance Analytics Emotion Analysis Motivation and Engagement	In addition to the goals of traditional training: Improve training effectiveness psychomotor and Procedural skills Spatial awareness Presence,immersion and realism Motion sickness Improve Design and Usability Enhance User Interface (UI) Optimize User Experience (UX) AI-Assisted Training
When	During Training Sessions After completing the task Hybrid scenario: During and after training	During Virtual Training Sessions After completing the task Hybrid scenario: During and after training
How	Participants verbalize thoughts during tasks Record and analyze verbalization	Participants verbalize thoughts while interacting with virtual systems Record and analyze verbalization

What In traditional training, TAP is considered a versatile tool that evaluates cognitive and intrapersonal skills, decision-making, and problem-solving (Zhang & Zhang, 2019). TAP is applied widely in education (Ciccarelli, 2024), skills enhancement (Jing et al., 2023), psychology, and human factors research (Jääskeläinen, 2012). In virtual training, TAP extends its applications to include Human-Computer Interaction (HCI) (Marco-Ruiz et al., 2017), usability testing (Hertzum, 2024), and simulation (Afifi et al., 2022) design (Negrão et al., 2023) research. It also supports training and simulation in virtual environments, making it a valuable method for understanding user interactions with digital systems such as websites, apps, VR/AR systems, wearable devices, and the global metaverse. This dual applicability highlights TAP's adaptability across both physical and virtual domains.

Who TAP utilizes diverse users in traditional and virtual training. In traditional settings, students, trainees, instructors, and professionals use TAP to uncover cognitive processes and enhance training outcomes. In virtual settings, designers, developers, researchers, and end-users apply TAP to interact with digital systems or virtual environments, highlighting its applicability across education, psychology, technology, and design.

Where The application of TAP varies depending on the context of the training. In traditional training, TAP is used in classrooms, workshops, and laboratories, where participants engage in hands-on activities and problem-solving tasks. In virtual training, TAP is employed in digital environments like websites, apps, VR/AR systems, wearable devices, and the metaverse, enabling evaluation of user interactions and experiences in immersive, interactive platforms, thus broadening TAP's applicability.

Why Numerous training and education applications emerge daily, yet many fail due to inadequate assessment methods. The primary goal of TAP in traditional training is to improve education and training programs' assessment procedures, improve cognitive and intrapersonal skills, and support forensic psychology and behavioral analysis. It also assesses managing cognitive load, analyzing performance, and understanding emotions, motivation, and engagement. This research expands TAP utilization opportunities, especially in virtual training where TAP-Express can be utilized to address additional objectives during the evaluation procedures, such as improving training effectiveness (Stefan et al., 2023), enhancing psychomotor and procedural skills, improving spatial awareness (Radhakrishnan et al., 2021), and addressing challenges like motion sickness (Saredakis et al., 2020) and its causes and measurements (Chang et al., 2020). TAP-Express directly improves user interfaces (UI) and enhances user experience (UX) (Kim et al., 2020), supporting AI-assisted training, which makes it essential for advancing virtual training programs and digital system design. In sport domains, TAP additionally promotes metacognition, reflection-in-action, and self-regulation (Birch et al., 2022; Stodter & Whitehead, 2024), complementing cognitive load management and performance analytics.

When TAP's application was initially utilized during the sessions that capture real-time cognition (Ericsson & Simon, 1993); on the other hand, according to Van Gog et al, TAP was utilized in post-task, in order to reduce cognitive load, but risks recall bias (Van Den Haak et al., 2003). Traditional settings use TAP in classrooms or labs for hands-on tasks. Digital platforms like apps, VR/AR, or the metaverse employ TAP to assess user interactions, broadening its scope.

How TAP records verbalized thoughts during tasks to uncover cognitive processes. In traditional training, participants articulate thoughts during or after physical tasks. In virtual training, they verbalize while engaging with digital systems, informing system design, usability, and user experience. This consistent methodology ensures TAP's reliability and adaptability across both training approaches (Eccles & Arsal, 2017).

This framework underscores the adaptability of TAP as a tool for training evaluation and sets the stage for the proposed TAP-Express method, which simplifies and streamlines the process for broader applicability.

4. TAP-express steps for training evaluation

TAP-Express is designed with philosophical flexibility, allowing alignment with positivist, interpretivist, or pragmatist approaches depending on the research goals. For instance, under interpretivism, verbalizations can be analyzed for subjective interpretations, while pragmatism enables integration of qualitative insights with quantitative success metrics to address real-world training challenges.

TAP-Express is a methodological framework for assessing cognitive processes in training, suitable for social sciences such as education (Ciccarelli, 2024) and healthcare (Zhao et al., 2024). It offers a standardized, replicable protocol that overcomes traditional TAP's limitations, ensuring effective application in both traditional and virtual environments. The eight steps below detail its implementation.

4.1. Step 1: define the training task

Clearly outline the training task that participants will perform. Ensure that the task is representative of the cognitive processes as shown in Table 2. The task design is informed by Yang et al. (2024), which provides evidence for predefining decision points in VR tasks to reduce cognitive load. Success criteria are explicitly defined, with purposive sampling to ensure representativeness (Cohen, 2013).

The training task is specified, focusing on critical decision points. Researchers predefine these points through task analysis to reduce participant cognitive load. For example, in a VR safety training task, critical moments include selecting a fire extinguisher type for a simulated chemical fire, assessing knowledge of safety protocols, and decision-making skills (Yang et al., 2024). For replicability, tasks are documented with objectives, procedures, and success criteria. Another example is a VR surgical simulation in which trainees perform a laparoscopic incision, assessing technical skills and anatomical knowledge, with success measured by incision accuracy.

These examples illustrate TAP-Express's applicability: the VR task involves 3–5 trainees verbalizing during a 10-minute simulation, tracked via VR logs. It assesses specific knowledge (e.g., surgical anatomy) and skills (e.g., precision), ensuring tasks have clear success levels. Participants are selected via purposive sampling to ensure representativeness, with inclusion/exclusion criteria defined to align with training objectives and enhance reproducibility.

4.2. Step 2: define variables and metrics

Before analysis, define variables and metrics aligned with the training context, either traditional or virtual, and evaluation goals. These variables provide a framework for assessing performance, engagement, usability, and other key training aspects. Table 3 and Fig. 3 summarizes key variables and their purposes based on the 5W1H framework. This list is a starting point, and researchers can select from it.

4.3. Step 3: instruct participants

Before beginning the TAP-Express procedure, participants must be properly instructed to ensure accurate and natural verbalization as shown in Table 4. This step includes clear instructions to optimize their performance without increasing cognitive strain. Structured guidance allows participants to understand when and how to verbalize their thoughts. The instructions should set expectations, reduce hesitation, and reinforce key decision points as the moments for verbalization.

Table 5 provides participants with guidelines copy to follow the TAP-Express process, ensuring clear and natural verbalization at key moments.

4.4. Step 4: warm-up and pilot evaluation

A warm-up procedure assists participants to become comfortable with the TAP-Express method as shown in Table 6. It reduces hesitation and ensures that participants understand how to verbalize their thoughts naturally.

Participants practice verbalization with a warm-up task. A neutral task is a low-stakes activity unrelated to training content, for instance, arranging geometric shapes in a 5-minute puzzle, or applying part of the original task, designed to develop verbalization skills and knowledge of the think-aloud process. While not entirely neutral due to potential familiarity effects, it minimizes bias, as supported by Eccles and Arsal (2017). This reduces hesitation and ensures consistent verbalization. For reproducibility, the task is standardized, for example, the same puzzle for all participants, with a 5-minute duration and written instructions provided.

Table 2
Framework for defining the training task.

Aspect	Description
Task	Clearly describe the task participants will perform, ensuring it reflects the cognitive processes being studied
Goal	Define the specific learning or performance outcomes and measurable success criteria
Complexity	Set the task difficulty (simple, moderate, complex) based on the cognitive effort required and participant expertise
Environment	Specify the environment (e.g., classroom, VR) and any factors that may affect performance
Resources	List tools, materials, or technologies needed (e.g., books, VR headsets, software, guides)
Metrics	Define Variables and Metrics (e.g., UX, cognitive load) (Step 2 and 6) Define success measures (e.g., accuracy, time, error rate) and ensure consistent evaluation (Step 7)
Instructions	Provide clear, step-by-step guidance to participants on task execution and expectations. (Step 3) with a warm-up procedure (Step 4)
Data Collection	Specify how participant responses will be recorded (e.g., verbal protocols, logs, physiological data). (Step 5)
Key Decision Points	Identify key moments such as decision-making, problem-solving, and information processing. Ensuring effective cognitive analysis while minimizing cognitive load. (Step 8)

Table 3
Key training variables categorized by 5W1H and training.

Category	ID	Variable Name	Training
What (Training Focus)	V1	Cognitive and Intrapersonal Skills	Both
	V2	Decision-Making and Problem-Solving	Both
	V3	Usability Testing	Virtual
	V4	Psychomotor and Procedural Skills	Virtual
	V5	Spatial Awareness	Virtual
	V6	Human-Computer Interaction (HCI)	Virtual
Who (Participants)	V7	User Engagement	Both
	V8	Instructor Guidance	Traditional
	V9	AI-Assisted Training	Both
Where (Environment)	V10	Training Effectiveness	Both
	V11	Classroom and Laboratory Performance	Traditional
	V12	Presence Immersion and Realism	Virtual
	V13	Metaverse and VR-based Learning	Virtual
Why (Purpose)	V14	Knowledge Retention	Both
	V15	Cognitive Load	Both
	V16	Behavioral and Psychological Analysis	Both
	V17	Motion Sickness	Virtual
	V18	Motivation and Engagement	Both
When (Timing)	V19	Real-Time Training	Both
How (Method)	V20	Verbalization Process	Both
	V21	UX/UI Design Evaluation	Virtual
	V22	AI-Based Personalized Feedback	Virtual

4.5. Step 5: conduct the session

Have trainees perform the task while verbalizing their thoughts at the specified points. Record the session for later analysis. Ensure a systematic approach to data collection and annotation to align it with observable performance metrics. To control for bias, standardize environmental factors such as noise, equipment, and participant characteristics, for instance, prior experience across sessions, logging these variables to ensure consistent data collection. Verbalizations are timestamped and annotated with categories to align with performance metrics, with controls maintaining uniform task conditions (Yang et al., 2024). A session protocol manual outlines and logs control procedures for replicability.

The following steps guide the process:

- Timestamp trainees' verbalizations at key moments, e.g., decision-making, problem-solving, and align with task progression logs for precise tracking and accurate analysis.

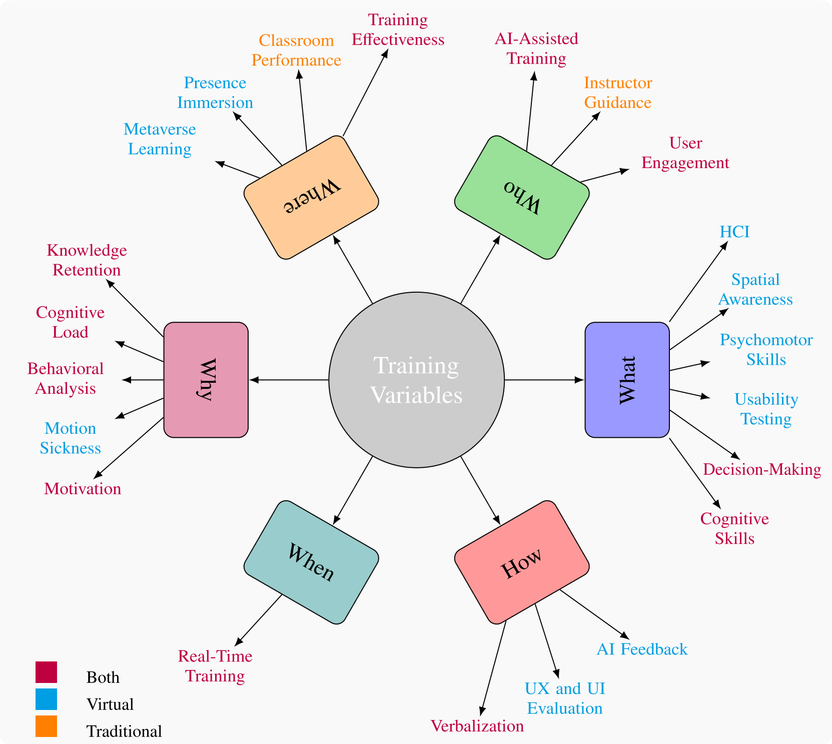


Fig. 3. 5W1H framework for training variables with color-coded training types.

Table 4
TAP-express participant instructions.

Instruction	Guideline
Purpose	Verbalize thoughts at key moments to help analyze decision-making.
When to Speak	Only at decision points, challenges, or critical observations.
What to Say	Speak your thoughts naturally without explaining or filtering.
How to Speak	Imagine talking to yourself; do not plan or structure responses.
Avoiding Silence	Say whatever comes to mind while silence reduces effectiveness.
Researcher's Role	Observers will not interfere but may remind you to verbalize.

Table 5
TAP-express participant guide.

Step	What You Need to Do
1. Speak Your Thoughts	Say whatever comes to mind at key moments (Tasks, decisions, challenges, observations).
2. No Need to Explain	Just verbalize your thoughts, you don't need to justify or analyze them.
3. Talk Naturally	Imagine talking to yourself; don't plan or organize your words.
4. Avoid Long Silences	If unsure, say what you are thinking, even if it seems unimportant.
5. Researcher's Role	The researcher will listen but not help. If you stop speaking, they may remind you.

Table 6
Warm-up and pilot evaluation for TAP-express.

Phase	Description
Warm-Up Task	Participants complete a neutral task before training to practice verbalization.
Instructions	Help participants adjust to speaking at key decision points without over-explaining.
Pilot Evaluation	A small-group test session evaluates participants' ability to verbalize naturally at decision points.
Assessment Criteria	Comfort Level: Are participants speaking naturally at key moments? Verbalization Accuracy: Do they verbalize at decision points enough? Cognitive Load Balance: Is verbalization affecting task performance?
Adjustments if Needed	If participants struggle, reinforce the warm-up exercise, adjust instruction clarity, and encourage shorter, more spontaneous responses.

- Annotate in Real-Time: Use observers or tools to label verbalized thoughts as they happen, e.g., “decision-making” or “problem-solving”. This organizes qualitative data into clear categories during the session.
- Record Surrounding Variables: Log environmental factors and technical issues affecting trainees, to capture all influences on the training process.

In addition to timestamping and annotating verbalizations for quantitative analysis, researchers may employ qualitative methods, such as thematic analysis or interpretive phenomenological analysis (as discussed in Section 2.1), to explore deeper patterns in participants' verbal reports. This involves transcribing recordings and coding for themes like cognitive strategies or subjective experiences, ensuring a holistic evaluation that complements the performance metrics computed in subsequent steps.

4.6. Step 6: compute TAP performance

The novelty of TAP-Express was in converting qualitative TAP data into a quantitative metric that enables rigorous analysis, streamlined performance comparisons, and actionable insights for training improvement. Researchers calculate overall performance by integrating key training variables with weighted adjustments, quantifying training effectiveness. This approach defines a performance formula using measurable variables (V1, V2, …) from the Key Training Variables (Table 3).

The weighted summation method integrates both positive and negative variables to evaluate training performance, wherein positive variables contribute directly to the performance score, while negative variables, such as motion sickness or cognitive load, are subtracted from 100 to reflect their inverse impact. The final score is calculated as:

$$P = \left(\sum_{i \in \mathcal{P}} w_i V_i + \sum_{j \in \mathcal{N}} w_j (100 - V_j) \right) \tag{1}$$

Where:

- P is the final performance score.
- w_i, w_j are the assigned weights for each variable.
- V_i represents positive variables ($\mathcal{P}$).
- V_j represents negative variables ($\mathcal{N}$), which are subtracted from 100 to normalize their impact.

The formula serves as a flexible template for further studies, as it guarantees that both positive and negative factors have been appropriately weighted.

TAP Performance employs an ordinal scale to rank verbalization quality (low, medium, high) based on clarity and relevance (Stevens, 1946). To address measurement theory concerns, non-parametric methods (e.g., Kruskal-Wallis test) are employed for group comparisons, ensuring valid statistical inferences in educational settings (Siegel & Castellan, 1988). Inter-rater reliability is calculated to validate scores, enhancing reproducibility (Field, 2013). Standardized scoring templates are provided to ensure consistency across studies.

Calculating performance using selected variables from Table 3, assigning weights based on training objectives. Statistical reporting includes confidence intervals, e.g., 95% CIs via bootstrapping and effect sizes, e.g., Cohen's d to quantify performance differences. Sensitivity analysis on weights ensures robust results across variable configurations. A performance calculation template with statistical protocols supports reproducibility.

4.6.1. Example calculation

Table 7 presents an example of training evaluation variables.
Using the formula, we compute P:

$$P = (0.3 \times 80) + (0.2 \times 85) + (0.25 \times (100 - 60)) + (0.25 \times (100 - 45))$$

$$P = 24 + 17 + 10 + 13.75 = 64.75\%$$

This approach allows future researchers to adjust variables and weights based on their specific training objectives, providing a structured yet adaptable performance evaluation model.

4.7. Step 7: compute success rate measurement

Success rate measurement quantifies training effectiveness by evaluating participants' ability to complete tasks efficiently and accurately. Unlike the TAP, the success rate analysis provides a structured observational assessment. The process involves tracking key performance indicators that reflect task completion, accuracy, efficiency, and corrective actions.

The proposed success rate model incorporates the following five variables, as summarized in Table 8.

Success Rate Measurement quantifies performance based on Step 1 criteria. This metric provides quantitative data, for instance, the percentage of correct decisions to complement qualitative verbalizations, enhancing evaluation via mixed methods. For the classroom task, success is the percentage of groups reaching consensus (e.g., 80%); for the VR task, it's the percentage of accurate incisions (e.g., 90%). Criteria are predefined to ensure measurable outcomes. The Success Rate is computed using a percentage scale (0–100%), justified as a ratio scale suitable for arithmetic operations (Stevens, 1946). To address the scale assumptions, a sensitivity analysis validates the metric's robustness across training contexts (e.g., VR simulations). Clear success criteria are predefined in Step 1 to ensure meaningful interpretation. A standardized protocol for data collection (e.g., observer logs, VR analytics) and statistical reporting supports replicability.

Table 7
Example variables and weight assignments.

Variable	Type	Value (%)	Weight
Decision-Making Efficiency (V_2)	Positive	80	0.3
Motivation (V_{18})	Positive	85	0.2
Motion Sickness (V_4)	Negative	60	0.25
Cognitive Load (V_{15})	Negative	45	0.25

Table 8
Success rate measurement variables.

Variable	Definition
SV_1	Task Completion Rate (%)
SV_2	Accuracy in Decision Making (%)
SV_3	Correction Success (%)
SV_4	Error Frequency (%)
SV_5	Time Efficiency (%)

Computing the success score using variables from Table 8, with weights reflecting task priorities. Statistical reporting includes 95% CIs via bootstrapping and effect sizes to assess performance robustness. Sensitivity analysis on weights validates the formula's stability across contexts. A success rate template with statistical reporting protocols ensures replicability.

Variable Transformation to Percentage:

To standardize the evaluation, raw data must be converted into percentages. Each variable is normalized based on its maximum possible value. The general transformation formula is:

$$SV_i = 100 \times \frac{\text{Observed Value}}{\text{Maximum Possible Value}} \tag{2}$$

For negatively impacting variables, such as error frequency, the formula is adjusted:

$$SV_j = 100 \times \left(1 - \frac{\text{Observed Value}}{\text{Maximum Possible Value}}\right) \tag{3}$$

Where SV_i represents positive metrics (e.g., completion, corrections), and SV_j represents negative metrics that negatively affect success scores. The overall success score (S) is computed using a weighted summation approach:

$$S = \sum_{i \in \mathcal{P}} w_i SV_i + \sum_{j \in \mathcal{N}} w_j (100 - SV_j) \tag{4}$$

Where:

- S is the overall success score.
- SV_i represents positive performance variables ($\mathcal{P}$).
- SV_j represents negative performance variables ($\mathcal{N}$), which are subtracted from 100 to reflect their inverse impact.
- w_i, w_j are the respective weight coefficients, which can be adjusted based on research needs.

4.7.1. Example calculation

First, we calculate the weight of the variables.

- **Task Completion Rate (SV_1):** In case of participant completed 4 out of 5 tasks, the percentage is:

$$SV_1 = 100 \times \frac{\text{Tasks Completed}}{\text{Total Tasks}} = 100 \times \frac{4}{5} = 80\%$$

- **Accuracy in Decision-Making (SV_2):** In case of participant has 17 correct decisions out of 20 total:

$$SV_2 = 100 \times \frac{\text{Correct Decisions}}{\text{Total Decisions}} = 100 \times \frac{17}{20} = 85\%$$

- **Correction Success (SV_3):** In case of participant corrects 6 out of 8 mistakes:

$$SV_3 = 100 \times \frac{\text{Successful Corrections}}{\text{Total Errors}} = 100 \times \frac{6}{8} = 75\%$$

- **Error Frequency (SV_4):** In case of the participant made 4 errors out of a maximum of 10, the percentage is computed as:

$$SV_4 = 100 \times \frac{\text{Errors Made}}{\text{Max Errors}} = 100 \times \frac{4}{10} = 40\%$$

- **Time Efficiency (SV_5):** In case the maximum time was 30 minutes and the participant completed in 18 minutes:

$$SV_5 = 100 \times \frac{\text{Max Time} - \text{Actual Time}}{\text{Max Time}} = 100 \times \frac{30 - 18}{30} = 40\%$$

Table 9 provides calculation with assigned weights for the example. The total success score is calculated as follows:

$$S = (0.30 \times 80) + (0.25 \times 85) + (0.20 \times 75) + (0.15 \times (100 - 40)) + (0.10 \times 40)$$

$$S = 24 + 21.25 + 15 + 9 + 4 = 73.25\%$$

Utilizing success score measurement provides a flexible template for evaluating success rates in different training scenarios, allowing researchers to adjust weights and variables based on their experiments.

4.8. Step 8: provide feedback

Transform the numbers in TAP-Express and success score measurement into useful information. TAP-Express analysis reveals training limitations and opportunities for improvement.

4.9. Pilot study design for TAP-express

To facilitate TAP-Express utilization, this research proposes a pilot study design as an example that researchers can adapt to test the effectiveness of the protocol in various training contexts. This pilot study provides a reproducible framework for TAP-Express utilization, aligning with the need for accessible methodologies in the social sciences. In a virtual training context, such as a surgical simulation task, with a small sample size, where trainees navigate a virtual environment, verbalizing decisions, e.g., applying an incision task. Success metrics include error frequency and time efficiency, while verbalization is coded for cognitive processes like Psychomotor and Procedural Skills (Yang et al., 2024). The pilot study design ensures TAP-Express's applicability to technological training, supporting interdisciplinary research.

For further analysis, this pilot study selected the surgical simulation task as a focal point. The following extended hypothetical pilot study about the surgical simulation was assessed by TAP-Express. All eight steps are covered in the pilot study, which gives researchers a thorough, reproducible template. To improve comprehension To improve clarity, Fig. 4 includes a flowchart that shows the main tasks and flow of each step.

A VR-based surgical simulation example evaluates decision clarity (Step 6) and task completion accuracy (Step 7), using standardized audio recordings and observer annotations for replicability. The pilot study includes research questions (e.g., Does TAP-Express effectively capture cognitive processes?) and tools like task definition tables. A statistical review is recommended to validate scale assumptions, ensuring rigor.

Table 9
Example success rate calculation.

Variable	Value (%)	Weight
Task Completion Rate (SV_1)	80	0.30
Accuracy in Decision-Making (SV_2)	85	0.25
Correction Success (SV_3)	75	0.20
Error Frequency (SV_4)	40	0.15
Time Efficiency (SV_5)	40	0.10

In the VR surgical simulation task, trainees perform a laparoscopic incision, assessing anatomical knowledge and technical skills. Critical decision points, e.g., selecting an incision site, are predefined via task analysis to reduce cognitive load (Yang et al., 2024). This example ensures clear success criteria, e.g., 90% incision accuracy within 2 mm, addressing replicability for diverse researchers.

To ensure reproducibility and accessibility, the pilot study template design is for 10–15 participants, e.g., teachers or medical trainees, in a controlled setting, e.g., a classroom or VR lab. This template provides a standardized and replicable protocol that is accessible to researchers and training developers. The study evaluates a training task using the eight TAP-Express steps, with detailed procedures, data collection methods, and analysis techniques to ensure methodological rigor. Fig. 4 visualizes the workflow, guiding researchers through each step's activities. This guide helps researchers in replicating the protocol. A statistical review of weighted formulas and statistical outputs is recommended to validate robustness.

1. **Step 1: Define the Training Task**
   To clarify the training task, training elements should be identified, such as the task type, the trainee sample involved, and the training method, in this case, a VR surgical simulation task where selected medical trainees perform a 10-minute laparoscopic incision. The task is documented with objectives, e.g., demonstrate anatomical knowledge, procedures, e.g., use VR scalpel to make an incision, and success criteria, e.g., accuracy within 2 mm. Key decision points, such as choosing the incision site or adjusting scalpel angle, are predefined using task analysis to minimize cognitive load, as supported by Yang et al. (2024). The task assesses knowledge, e.g., heart anatomy, and skills, e.g., precision, with VR logs capturing performance data. A task definition table, e.g., Table 2 is provided to ensure replicability across studies. Participants are selected via purposive sampling, targeting trainees with 1–3 years of surgical experience, with inclusion criteria, e.g., VR familiarity, and exclusion criteria, e.g., fatigue, prior task exposure to ensure representativeness (Cohen, 2013).
2. **Step 2: Define Variables and Metrics**
   Defining variables from Table 3, such as V2: Decision-Making, positive, and V15: Cognitive Load, negative, measured via verbalizations and VR analytics, e.g., hesitation time. Weights are assigned, e.g., 0.4 for V2, 0.3 for V15, to compute performance Eq. (1). For the VR task, decision-making accuracy e.g., correct incision site choice, and cognitive load, e.g., verbalized hesitation, are quantified, ensuring measurable outcomes aligned with training goals (Zhang & Zhang, 2019). A variable selection checklist is included in the manual for reproducibility.
3. **Step 3: Instruct Participants**
   Providing written instructions as in Table 4 and a participant guide as in Table 5 for 10–15 trainees, explaining verbalization at key decision points, e.g., "Say why you chose this incision site". A 5-minute briefing ensures understanding. Instructions emphasize natural verbalization, e.g., "I'm choosing this site because it's near the heart" to reduce cognitive strain, with reminders from observers if silence occurs (Eccles & Arsal, 2017). The standardized briefing script ensures consistency across studies.
4. **Step 4: Warm-Up and Pilot Evaluation**
   Conducting a 5-minute warm-up with a neutral VR navigation task, e.g., finding a virtual exit in a neutral environment, standardized for all participants with written instructions. This task develops verbalization skills, e.g., "I'm turning right because the path is open" and knowledge of think-aloud, unrelated to surgical content, to minimize bias (Eccles & Arsal, 2017). A pilot evaluation with 5 participants assesses verbalization comfort

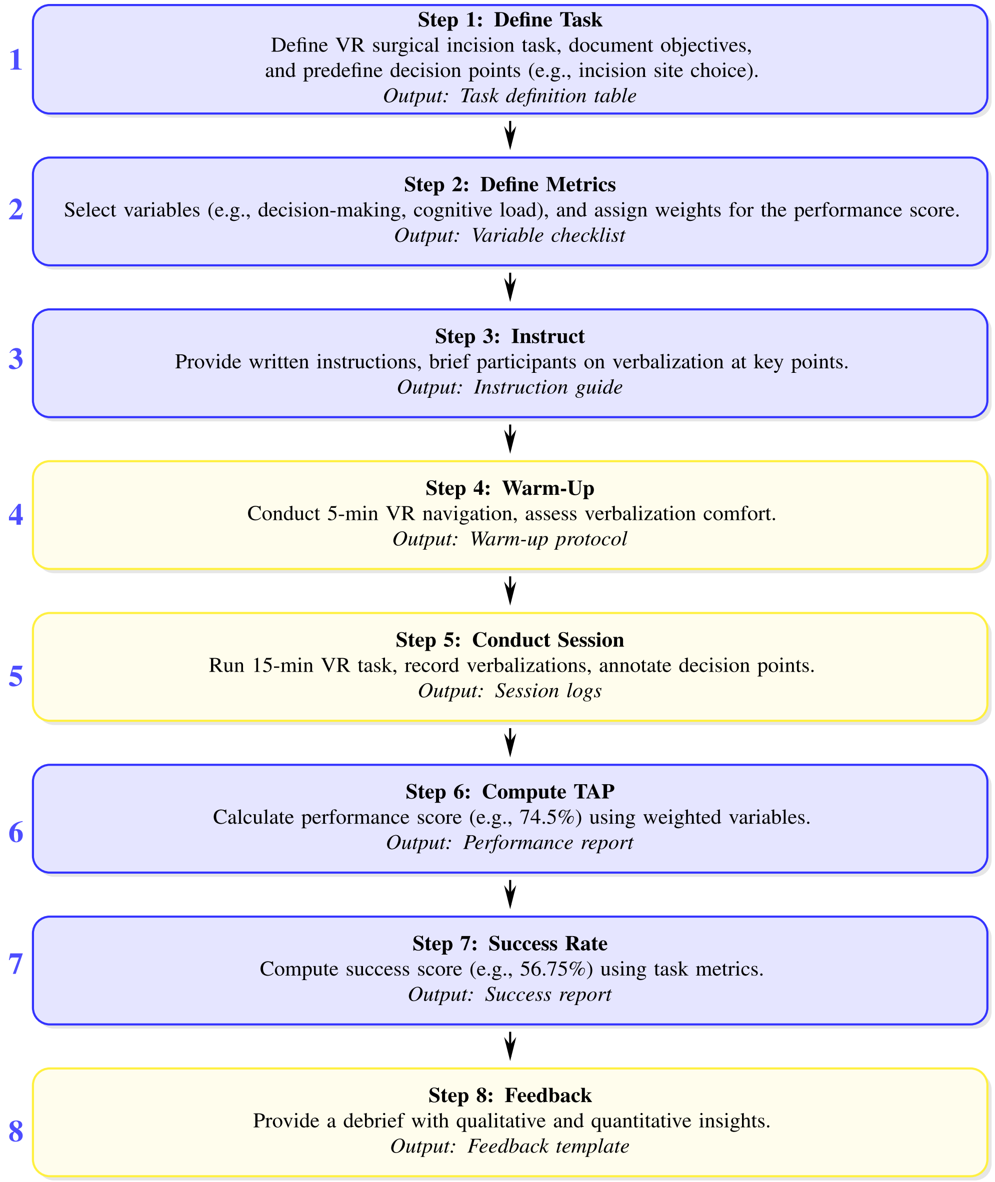


Fig. 4. TAP-express pilot study steps: key activities and outputs.

and adjusts instructions if needed. Audio recordings and observer notes ensure reproducible assessment criteria as in Table 6 procedures.

5. **Step 5: Conduct the Session**

   Running a 15-minute VR surgical task session for 10–15 trainees, with audio/video recording of verbalizations and VR logs for performance data. Verbalizations are timestamped at key moments, e.g., incision site selection, annotated as “decision-making” or “problem-solving,” and linked to VR actions, e.g., scalpel movement. Environmental factors, e.g., lab noise, are logged to contextualize data. A session protocol manual standardizes data collection for replicability. Environmental factors, e.g., lab noise, VR headset model, UX, and participant characteristics like prior VR experience, with variables logged to minimize bias.

6. **Step 6: Compute TAP Performance**

   Calculating performance using Eq. (1), combining V2 (decision-making, 80% score) and V15 (cognitive load, 45% score) with weights, e.g., P = 0.4 × 90% + 0.7 × (100 - 45)%. For example, a trainee’s verbalizations show 80% correct decisions and moderate cognitive load, yielding a performance score of 74.5%, providing quantitative insights into training effectiveness. A performance calculation template is included for other researchers. Performance score includes a 95% confidence interval (CI) estimated via bootstrap resampling, with sensitivity analysis (e.g., varying the weight of key indicators) showing consistent results within a moderate performance range, and an effect size (e.g., Cohen’s d indicating a medium effect) to ensure robustness.

7. **Step 7: Compute Success Rate Measurement**

   Computing the success score using Eq. (4), with variables from Table 8, e.g., SV1: Task Completion, 90%; SV2: Accuracy, 85% and weights, e.g., 0.3 for SV1, 025 for SV2. For the VR task, success is 90% incision accuracy, calculated as S = 0.3 × 90 + 0.35 × 85, yielding 56.75%. A success rate template with sample calculations as in Table 9 to ensure reproducibility. Success score includes, for instance, 95% CI via 1000 bootstrap iterations, with sensitivity analysis (e.g., V1 weight range) and effect size (Cohen's d = 0.8) to validate results.

8. **Step 8: Provide Feedback**
   Analyzing qualitative, e.g., verbalized decision strategies and quantitative, e.g., 74.5% performance score, 56.75% success score data to provide feedback, by identifying strengths, e.g., accurate incision placement and areas for improvement, e.g., hesitation in site selection. Feedback is delivered via a 10-minute debrief, linking verbalizations to performance, e.g., "Your hesitation at 3:12 suggests uncertainty in heart anatomy" to guide training revisions. A feedback template ensures consistent reporting for future studies.

This pilot study template, supported by a standardized manual and Fig. 4 along with Fig. 5 for statistical guidelines for a general overview of variables, formulas, and statistical reporting to guide researchers in replicating the analysis. This ensures TAP-Express's replicability across social science applications, supporting diverse researchers by providing clear, adaptable procedures.

Addressing the three research questions: first, evaluating TAP-Express's effectiveness; second, its applicability; and third, its reproducibility in training evaluation. A pilot study yielding a 74.5% TAP Performance score and 56.75% Success Rate in a VR surgical simulation suggests TAP-Express streamlines the TAP, tackling resource-intensive constraints (Noushad et al., 2024), pending further validation. TAP-Express targets key decision moments, such as incision site selection in a VR surgical simulation, reducing participant burden while providing clearer insights compared to traditional TAP (Ericsson & Simon, 1993), which leads to streamlined cognitive process capture. It is applicable to both traditional (e.g., classroom tasks like group consensus) and virtual (e.g., surgical simulations) training contexts, offering versatility (Zhao et al., 2024). Additionally, the reproducibility of TAP-Express in social sciences, such as education and healthcare, through standardized protocols and a pilot study template (Yang et al., 2024). Non-parametric methods, like the Kruskal-Wallis test, sensitivity analyses, and Cohen's d effect size, ensure statistical robustness. TAP-Express blends qualitative and quantitative methods for education and psychology (Ciccarelli, 2024), where it requires larger-scale testing by targeting critical decision points and measuring success rates. The pilot study's standardized protocols provide a replicable framework, ensuring accessibility and rigor for diverse researchers.

TAP-Express's flexibility supports traditional and virtual training contexts for mitigating barriers like setup time and analysis complexity (Ericsson & Simon, 1993), and its worked examples serve as a methodological template for social science research.

5. Discussion

In this section, TAP-Express discusses its features by analyzing the research questions answers, starting with the evaluation of the effectiveness of TAP-Express, and then going through the applicability of TAP-Express, and finally the reproducibility of TAP-Express in training evaluation. The pilot study results show that TAP Performance has a high score and moderate success rate in a VR surgical simulation, thus suggesting TAP-Express streamlines the TAP, tackling resource-intensive constraints (Noushad et al., 2024), pending further validation. TAP-Express streamlines traditional TAP, addressing resource-intensive constraints (Noushad et al., 2024). First, TAP-Express effectiveness lies in capturing cognitive processes by focusing verbalizations on critical decision points, such as incision site selection in a VR surgical simulation, reducing participant burden while providing clear insights into decision-making compared to traditional TAP's broader approach (Ericsson & Simon, 1993). It merges qualitative cognitive insights with quantitative success metrics, improving applicability in social sciences such as education and psychology (Ciccarelli, 2024). Second, TAP-Express can be utilized in both non-technological (traditional) and technological (virtual) training contexts, as demonstrated by its use in classroom tasks, for example, group consensus and VR simulations such as surgical training, offering a versatile methodology for diverse settings (Zhao et al., 2024). Statistical robustness is ensured through non-parametric methods, for instance, the Kruskal-Wallis test, sensitivity analyses, and effect

Statistical Summary for TAP-Express
Variables, Formulas, and Reporting

Variables
TAP Performance: e.g., Decision-Making, Cognitive Load
Success Rate: e.g., Task Completion, Accuracy, Correction Success
Source: Tables 3, 8

Formulas
TAP Performance: $P = \sum_{i\in\mathcal{P}} w_i V_i + \sum_{j\in N} w_j(100 - V_j)$
Success Rate: $S = \sum_{i\in\mathcal{P}} w_i S V_i + \sum_{j\in N} w_j(100 - S V_j)$
Source: Equations 1, 4

Statistical Reporting
95% CIs through bootstrapping
Effect Sizes (e.g., Cohen's d)
Sensitivity Analysis for weight variations
Reproducible through templates in Sections 4.6, 4.7

Fig. 5. Statistical summary for TAP-express.

size reporting, such as Cohen's d, addressing scale-level. A pilot study design with standardized protocols and a recommended statistician review ensures accessibility and rigor (Siegel & Castellan, 1988). Third, this research introduced a reproducible framework of TAP-Express for social science applications, such as education and healthcare, through standardized protocols and a detailed pilot study template, enabling researchers to replicate and adapt the methodology across contexts (Yang et al., 2024).

The protocol's steps show its versatility, and the pilot study paves the way for testing its value across various fields. Examples tied to training tasks reveal how TAP-Express tracks thinking processes, offering a practical guide for social science research. It also cuts down on setup time and simplifies analysis, easing common TAP challenges.

6. Strengths and contributions of TAP-express

In this section, TAP-Express's strengths were summarized to emphasize its simplicity, adaptability, and accessibility. Three main strengths of TAP-Express allow it to be more adept for social science research as follows: First, the simplicity of TAP-Express reduces the participants' verbalization requirement by focusing only on key decision points, making it practical for resource-constrained settings, such as classroom-based teacher training (Ciccarelli, 2024). Second, the adaptability of TAP-Express allows application across diverse contexts, including psychology and education (Eccles & Arsal, 2017). Third, the accessibility of TAP-Express supports researchers and underrepresented groups by providing a standardized, replicable protocol with minimal training requirements, as demonstrated in the pilot study. TAP-Express is positioned as a valuable tool for evaluating cognitive processes in training, contributing to accessible methodologies in social sciences and other fields. Fourth, overcoming traditional TAP's limitations by TAP-Express, such as time-intensive setup and complex data analysis (Ericsson & Simon, 1993; Van Den Haak et al., 2003), by offering a streamlined protocol that integrates Success Rate Measurement for rigorous quantitative evaluation, as shown in Section 4. Fifth, the flexibility of TAP-Express in virtual training, exemplified by the VR-based surgical simulation in the pilot study, addresses technological research gaps and improves the effectiveness of training in fields such as healthcare (Zhao et al., 2024) and other fields.

This research contributes to a robust methodological framework that supports interdisciplinary applications, aligning with the latest commitment to innovative and inclusive research.

7. Limitations

TAP-Express's shortcomings pave the way for new possibilities. While it simplifies evaluating training programs, its constraints point to areas needing work for solid use across varied settings. Key limitations include:

- **Lack of Empirical Validation**: TAP-Express remains a framework without empirical data to confirm its effectiveness. The pilot study design provides a template for future validation, but until tested, claims about its impact are limited (Ericsson & Simon, 1980).
- **Participant Verbalization Challenges**: The protocol relies on participants' ability to verbalize thoughts at key decision points, which may vary due to individual factors such as shyness or cognitive overload, potentially affecting data reliability (Eccles & Arsal, 2017).
- **Need for Deep Understanding of Model and Variables**: Effective implementation of TAP-Express requires researchers to have a thorough understanding of the protocol and its variables (e.g., those in Table 3), which may pose a challenge for those unfamiliar with think-aloud methods or training evaluation.

8. Conclusion

In conclusion, this research introduced the TAP-Express that offers a rigorous and practical solution for capturing and analyzing cognitive processes in diverse training environments. Answering the research questions, TAP-Express captures cognitive processes effectively through emphasizing verbalizations on critical decision points, such as incision site selection in virtual reality surgical simulations, providing clearer insights into decision-making compared to the traditional TAP's broader approach. It applies to both non-technological (traditional) and technological (virtual) training contexts, as shown in classroom tasks like group consensus and virtual tasks like surgical simulations, offering a versatile tool for diverse settings. Additionally, TAP-Express ensures reproducibility in social science applications, such as education and healthcare, through standardized protocols and a detailed pilot study template, enabling researchers to adapt and replicate the methodology across contexts. TAP-Express streamlines the TAP for evaluating training programs across traditional, virtual, online, and mixed-reality settings, targeting cognitive and intrapersonal skills. It simplifies verbalization while retaining cognitive insights and integrates success rate measurement for robust qualitative and quantitative evaluation. This structured, adaptable methodology, supported by the pilot study's reproducible design, provides a foundational template for diverse training contexts and future research. The eight-step framework improves methodological clarity and accessibility, particularly for early-career researchers. Integrating Success Rate Measurement links performance metrics with qualitative data, offering a balanced evaluation approach. Its reproducible structure, supported by standardized tools and pilot templates, ensures accessibility across disciplines. Despite challenges such as variability in participant verbalization and limited VR access, TAP-Express refines through iterative calibration to maintain rigor and adaptability. Designed for both traditional and immersive learning environments, TAP-Express supports inclusive and adaptive assessment strategies. This refined methodology provides a practical and transferable foundation for evaluating training programs and guiding future research. This streamlined, adaptable methodology integrates qualitative and quantitative evaluation, fostering accessible and innovative training assessments.

9. Future research directions

Future research can dive deeper into fixing TAP-Express shortcomings and expanding its applications across diverse settings. For example, further studies are needed for testing TAP-Express in large-scale virtual training environments, such as augmented reality or mixed reality platforms, which could make it more practical for real-world use. Additionally, cross-cultural studies might explore how well TAP-Express works in varied contexts, tackling challenges like language barriers or cultural differences in verbal expression. Long-term studies, spanning multiple years, could track its reliability across different training programs, ensuring consistent results over time. On the other hand, integrating automated tools, like AI-driven text analysis or real-time data processing software, could cut down on the time researchers spend sifting through verbal data, making the process smoother and less taxing. Additionally, experimenting with TAP-Express in emerging fields like virtual reality therapy or online soft-skills training could uncover new uses, while collaborations with interdisciplinary teams might refine its methodology. These efforts aim to strengthen training evaluation in social sciences, humanities, and beyond, making TAP-Express a go-to tool for understanding cognitive processes in diverse training scenarios.

CRediT authorship contribution statement

Abdallah Al-Hamad: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation,

Formal analysis, Data curation. **Attila Gilányi:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration.

Ethical statement

Ethical approval is not applicable to this manuscript.

Declaration of generative AI

AI tools were used to assist in language editing and proofreading, but the authors are fully responsible for the content.

Funding statement

This research received no external funding.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research has been supported by project no. 152101, implemented with the support provided by the Ministry of Culture and Innovation of Hungary from the National Research, Development and Innovation Fund, financed under the ADVANCED_25 funding scheme.

Supported by the University of Debrecen Program for Scientific Publication.

This research has been supported by project no. 2024-1.2.3-HU-RIZONT-2024-00030, implemented with the support provided by the Ministry of Culture and Innovation of Hungary from the National Research, Development and Innovation Fund, financed under the 2024-1.2.3-HU-RIZONT funding scheme.

Data availability

No data was generated for this research.

References

Abd Rahman, M. A., Salim, M. H. M., & Ali, N. M. (2023). The evaluation of a persuasive learning tool using a think-aloud protocol. *International Journal of Advanced Computer Science and Applications*, *14*. Available at: https://doi.org/10.14569/IJACSA.2023.0140635

Afifi, E. H. N., Diantoro, M., & Wartono, W. (2022). Analisis miskonsepsi siswa MA menggunakan metode think aloud protocol (TAP). *Jurnal Pendidikan: Teori, Penelitian, dan Pengembangan*, *7*, 72–79. Available at: https://doi.org/10.17977/jptpp.v7i2.15193

Byrd, N., Joseph, B., Gongora, G., & Sirota, M. (2023). Tell us what you really think: A think aloud protocol analysis of the verbal cognitive reflection test. *Journal of Intelligence*, *11*, 76. Available at: https://doi.org/10.3390/jintelligence11040076

Birch, P. D. J., Yeoman, B., & Whitehead, A. E. (2022). "Think Aloud" as a facilitator of self-regulation in golfers. *The Sport Psychologist*, *36*(4), 274–283. Available at: https://doi.org/10.1123/tsp.2022-0017

Chang, E., Kim, H. T., & Yoo, B. (2020). Virtual reality sickness: A review of causes and measurements. *International Journal of Human-Computer Interaction*, *36*, 1658–1682. Available at: https://doi.org/10.1080/10447318.2020.1778351

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York. Available at: Routledge. https://doi.org/10.4324/9780203771587

Ciccarelli, R. (2024). Comunicazione istituzionale e educazione linguistica democratica. Il contributo del think aloud protocol. *Italiano LinguaDue*, *16*, 85–97. Available at: https://doi.org/10.54103/2037-3597/23827

Eccles, D. W., & Arsal, G. (2017). The think-aloud method: What is it and how do i use IT? *Qualitative Research in Sport, Exercise and Health*, *9*, 514–531. Available at: https://doi.org/10.1080/2159676X.2017.1331501

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, *87*, 215–251. Available at: https://doi.org/10.1037/0033-295X.87.3.215

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (Rev. ed.). Cambridge, MA: Bradford Books/MIT Press. Original work published 1984.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). SAGE Publications Ltd.

Hertzum, M. (2024). Concurrent or retrospective thinking aloud in usability tests? A meta-analytic review. *ACM Transactions on Computer-Human Interaction.* Available at: https://doi.org/10.1145/3665327

Imbulpitiya, A., Whalley, J., & Senapathi, M. (2023). Reflections on conducting online think-aloud. *ACM Inroads*, *14*, 26–34. Available at: https://doi.org/10.1145/3596917

Jääskeläinen, R. (2012). Think-aloud protocol. In *Handbook of translation studies: Volume 1* (pp. 371–373). Available at: John Benjamins Publishing Company. https://doi.org/10.1075/hts.1.thi1

Jing, X., Draghi, B. N., Ernst, M. A., Patel, V. L., Cimino, J. J., Shubrook, J. H., Zhou, Y., Liu, C., & De Lacalle, S. (2023). How do clinical researchers generate data-driven scientific hypotheses? Cognitive events using a think-aloud protocol. *medRxiv*. Available at: https://doi.org/10.1101/2023.10.31.23297860

Kim, Y. M., Rhiu, I., & Yun, M. H. (2020). A systematic review of a virtual reality system from the perspective of user experience. *International Journal of Human-Computer Interaction*, *36*, 893–910. Available at: https://doi.org/10.1080/10447318.2019.1699746

Krishna-Naik, V., Palmer, A., Hodson, N. A., Tugnait, A., & O'Connor, D. B. (2024). Utilisation of a think-aloud protocol to validate a self-reported periodontitis questionnaire. *Journal of Dentistry*, *150*, 105381. Available at: https://doi.org/10.1016/j.jdent.2024.105381

Lewrick, M., Link, P., & Leifer, L. (2020). *The design thinking toolbox: A guide to mastering the most popular and valuable innovation methods*. John Wiley & Sons.

Marco-Ruiz, L., Bønes, E., de la Asunción, E., Gabarron, E., Aviles-Solis, J. C., Lee, E., Traver, V., Sato, K., & Bellika, J. G. (2017). Combining multivariate statistics and the think-aloud protocol to assess human-computer interaction barriers in symptom checkers. *Journal of Biomedical Informatics*, *74*, 104–122. Available at: https://doi.org/10.1016/j.jbi.2017.09.002

Marijuan, S. (2024). Think-aloud protocol, trials-to-criterion, and triangulation between online and offline measures in IP research. In *The routledge handbook of second language acquisition and input processing* (pp. 378–391). Available at: Routledge. https://doi.org/10.4324/9781003041412

Menon, R. K., Bhatia, S., Nadarajah, V. D., & Pau, A. (2022). Piloting online "Think Aloud" sessions to support clinical learning during the pandemic. *Education in Medicine Journal*, *14*. Available at: https://doi.org/10.21315/eimj2022.14.2.7

Negrão, M. D., Ferreira, W., Bohrer, B., Freitas, C. M., Maciel, A., & Nedel, L. (2023). Design and think-aloud study of an immersive interface for training health professionals in clinical skills. In *Proceedings of the 25th symposium on virtual and augmented reality* (pp. 157–165). Available at: https://doi.org/10.1145/3625008.3625037

Noushad, B., Van Gerven, P. W., & De Bruin, A. B. (2024). Twelve tips for applying the think-aloud method to capture cognitive processes. *Medical Teacher*, *46*, 892–897. Available at: https://doi.org/10.1080/0142159X.2023.2289847

Radhakrishnan, U., Koumaditis, K., & Chinello, F. (2021). A systematic review of immersive virtual reality for industrial skills training. *Behaviour & Information Technology*, *40*, 1310–1339. Available at: https://doi.org/10.1080/0144929X.2021.1954693

Saredakis, D., Szpak, A., Birckhead, B., Keage, H. A., Rizzo, A., & Loetscher, T. (2020). Factors associated with virtual reality sickness in head-mounted displays: A systematic review and meta-analysis. *Frontiers in Human Neuroscience*, *14*, 96. Available at: https://doi.org/10.3389/fnhum.2020.00096

Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). McGraw-Hill.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, *103*(2684), 677–680. Available at: https://doi.org/10.1126/science.103.2684.677

Stefan, H., Mortimer, M., & Horan, B. (2023). Evaluating the effectiveness of virtual reality for safety-relevant training: A systematic review. *Virtual Reality*, *27*, 2839–2869. Available at: https://doi.org/10.1007/s10055-023-00843-7

Stodter, A., & Whitehead, A. (2024). Thinking again about the use of think aloud and stimulated recall methods in sport coaching. *Qualitative Research in Sport, Exercise and Health*, *16*(5), 456–470. Available at: https://doi.org/10.1080/2159676X.2024.2377658

Van Den Haak, M., De Jong, M., & Jan Schellens, P. (2003). Retrospective VS. Concurrent think-aloud protocols: Testing the usability of an online library catalogue. *Behaviour & Information Technology*, *22*, 339–351. Available at: https://doi.org/10.1080/0044929031000

Yang, Z., Zhang, Z., Wu, W., & Xia, Z. (2024). A study on user's cognitive needs in virtual reality training systems from an embodied cognition perspective. In *2024 30th international conference on mechatronics and machine vision in practice (M2VIP)* (pp. 1–5). Available at: IEEE. https://doi.org/10.1109/M2VIP62491.2024.10746113

Zhang, L. J., & Zhang, D. (2019). Think-aloud protocols. In *The routledge handbook of research methods in applied linguistics* (pp. 302–311). Routledge.

Zhang, X., & Simeone, A. L. (2022). Using the think aloud protocol in an immersive virtual reality evaluation of a virtual twin. In *Proceedings of the 2022 ACM symposium on spatial user interaction* (pp. 1–8). Available at: https://doi.org/10.1145/3565970.3567706

Zhao, R., Ding, T., Meng, J., Lei, M., & Ma, H. (2024). Think-aloud approach combined with case-based learning in nursing teaching round for new nurses in cancer hospital. *BMC Medical Education*, *24*, 916. Available at: https://doi.org/10.1186/s12909-024-05891-4

Further reading

Badea, D. O., Darabont, D. C., Chis, T. V., & Trifu, A. (2024). The transformative impact of e-learning on workplace safety and health training in the industry: A comprehensive analysis of effectiveness, implementation, and future opportunities. *International Journal of Education and Information Technologies*, *18*, 105–118. Available at: https://doi.org/10.46300/9109.2024.18.11

Ericsson, K. A. (2017). Protocol analysis. *A Companion to Cognitive Science*, 425–432. Available at: https://doi.org/10.1002/9781405164535.ch33