American Journal of Education and Technology (AJET)

ISSN: 2832-9481 (ONLINE

VOLUME 5 ISSUE 3 (2026)

PUBLISHED BY
E-PALLI PUBLISHERS, DELAWARE, USA

American Journal of Education and Technology (AJET)

Volume 5 Issue 3, Year 2026
ISSN: 2832-9481 (Online)
DOI: https://doi.org/10.54536/ajet.v5i3.8173
https://journals.e-palli.com/home/index.php/ajet

Digital Exams as Socio-Technical Interaction Systems: An ISO 9241-Based Framework for Validity, Fairness and Governance

Giovanni Vindigni, D.Th.[1,2*]

Article Information

Received: August 17, 2025
Accepted: October 20, 2025
Published: August 01, 2026

Keywords

Accessibility, Assessment Validity, Computer-Based Assessment, Digital Exam Governance, DIN EN ISO 9241, Fairness, Human-Centred Design, Interaction Principles, Remote Proctoring

ABSTRACT

This paper develops a normatively explicit framework model for explaining and evaluating performance differences between handwritten and computer-based exams in tertiary education. It rejects the reduction of so-called mode effects to input modality and reconstructs digital exams as socio-technical examination systems whose interaction architecture can materially shape the conditions of valid and fair assessment. The study addresses a central gap in the literature: although performance differences between handwritten and computer-based formats have repeatedly been documented, they remain insufficiently explained in terms of the instructional, interactional, and contextual conditions under which digital examination systems operate. Anchored in DIN EN ISO 9241-11, -110, and -210, the study develops an explanatory and evaluative model and examines it against heterogeneous evidence from CBT-versus-paper comparability research, proctoring and integrity studies, writing-mode mechanisms, and stress- and strain-related findings. Selected results from the author's own work in digitally mediated tertiary settings are incorporated as contextual bridging arguments, without being reclassified as primary data. Across this reconstruction, performance differences are shown to be insufficiently explainable as modality artifacts alone and are more adequately understood as outcome effects of a socio-technical examination system in which context of use, system quality, error consequences, monitoring regimes, and accessibility conditions interact. In this logic, usability under DIN EN ISO 9241-11 is not an ancillary quality feature but a constitutive validity requirement. DIN EN ISO 9241-110 provides the operational bridge through which interface and process characteristics can be translated into exam-relevant fairness and risk profiles, while DIN EN ISO 9241-210 supplies the human-centered process perspective required for governance, lifecycle evaluation, and risk-adequate design adjustment. The contribution does not present a separate controlled usability experiment or a statistical meta-analysis; its epistemic contribution lies in reconstructing effect pathways, systematizing heterogeneous findings along ISO-based categories, and deriving testable propositions and minimum evaluative criteria. The resulting framework identifies interaction load, error exposure, monitoring regimes, and accessibility conditions as validity-relevant variables and offers a normatively grounded basis for the design, evaluation, and governance of digital exams. It thereby provides a theoretically robust and practically applicable foundation for universities seeking to implement digital examinations without reducing validity, fairness, or equitable participation to secondary considerations.

INTRODUCTION

Many view digital examinations in tertiary education primarily as a modernization of infrastructure, similar to the government's mandated digitization strategy. These improvements should be accompanied by faster distribution, more efficient assessment, better readability, and reduced logistical effort (Aini *et al.,* 2021; Chatzipetrou *et al.,* 2023; Coghlan *et al.,* 2021; Mahyoob, 2020; Raaheim *et al.,* 2019). However, this administrative rationality obscures at this point the fact that exam situations represent highly stressful, risky interactions in which even minor deviations in system behavior, error sequences, or contextual conditions can substantially influence the validity and fairness of performance assessments (Aini *et al.,* 2021; Bashir *et al.,* 2021; Browning *et al.,* 2021; Elmer *et al.,* 2020; Johnson *et al.,* 2020; Vindigni, 2025c).

In this context, the question of performance differences between handwritten and computer-based tests should therefore, not be treated as a marginal issue of input modality or literacy formation per se but as the subject of a socio-technical system analysis with regard to the ergonomics of human-system interaction (Deutsches Institut für Normung (DIN), 2018a, 2020b, 2020c).

In this regard, the central issue addressed in this paper is that while current research on digital exams in higher education has repeatedly documented performance differences between handwritten and computer-based formats, these differences are mostly treated heuristically as modality or medium effects, without sufficiently reconstructing the instructional design and socio-technical conditions under which digital exam systems are structured (Vindigni, 2026d). This approach overlooks

[1] Department of Media Business and Media Management, Digital Games Business, and UX-Design, Diploma Hochschule, University of Applied Sciences, Bad Sooden-Allendorf, Germany
[2] Department of Economics, Global Humanistic University, The Valley (AI). Doctoral Advisor/Doctoral Supervisor, Middlesex University (UK), UK
[*] Corresponding author's e-mail: giovanni.vindigni@diploma.de

the fact that digital exams are not merely a media-based adaptation of analog exam formats but effectively amount to independent interaction systems in which the context of use, quality of interaction, consequences of errors, control regimes, and accessibility conditions can jointly shape exam-related outcomes. A substantial desideratum lies precisely here, because observed performance differences remain theoretically underdetermined until it is explained whether they are due to differences in competence or to the construct-irrelevant effects of interaction, burden, and exclusion. The research that has dominated to date explains such differences predominantly in psychological, acceptance-related, or descriptive-comparative terms without systematically translating them into a normatively explicable framework of human-system interaction. This is all the more significant given that, in the European context, digital exams must be understood as exam-relevant infrastructures not only from the perspective of higher education pedagogy but also in terms of accessibility, equitable participation, and risk-appropriate governance (Vindigni, 2026d). This paper addresses precisely this gap by reconstructing digital exams based on DIN EN ISO 9241 not as a mere implementation issue but as a socio-technical examination system that can be evaluated normatively.

Reconstructing these interrelationships is not only theoretically warranted, but also necessary from a contemporary practical and governance standpoint, given the ongoing institutionalization of digital examination formats in higher education and their integration into binding requirements for accessibility and ensuring participation. The more digital exams become part of the standard examination infrastructure, the less their design can be treated as a mere question of implementation or efficiency; rather, it must be understood as a constitutive condition of validity, fairness, and equitable assessment of performance.

LITERATURE REVIEW

Phenomenological Initial Observation and Relevance

Phenomenologically, reports are accumulating in universities especially under pressure to digitize and following the pandemic-driven push toward digitization—that exam performance varies depending on the mode (handwritten vs. computer-based), even though the intended construct reference should remain identical (Ceka & O'Geen 2019; Van der Weel & Van der Meer 2023; Lestari 2024). This observation is not only didactically relevant but also touches on the core of institutional legitimacy of academic performance assessment: if mode conditions are systematically co-assessed, there is a risk that exams will no longer exclusively reflect competence but rather interaction management (including error, stress, and surveillance management) (Vindigni, 2025c). Here, the underlying premise that performance outcomes in digitized university settings do not arise in isolation from a difference in modality, nor from context-sensitive interaction, control, and stress effects, has been reconstructed in related digitally mediated teaching and learning settings based on ISO 9241 standards and is empirically compatible (Vindigni, 2025c, 2026b).

Mode effects pose a risk to validity, particularly in situations where extended response formats such as free text, justifications, and derivations are prevalent (Baragash & Al-Samarraie, 2018; Jones *et al.,* 2022; Wagner *et al.,* 2022; Wang *et al.,* 2008; Yanfei, 2021). A recent literature review on writing mode effects shows that comparability cannot be determined solely by score differences but also by marking practices, text characteristics, and composition processes—in other words, by mechanisms that are shaped in situ by the exam arrangement (Jones *et al.,* 2022; Khoshsima *et al.,* 2017; Lestari, 2024; Wagner *et al.,* 2022).

Therefore, the above context focuses on the thesis that digital exams in tertiary education should be understood as socio-technical interaction arrangements whose usability in the sense of DIN EN ISO 9241-11 represents a constitutive validity condition (Deutsches Institut für Normung (DIN), 2018a; Vindigni, 2025b, 2026c, 2026b). The article provides a normatively grounded system reconstruction that uses DIN EN ISO 9241-110 as an operationalization bridge to translate interface and process characteristics into exam-relevant fairness and outcome risks and anchors DIN EN ISO 9241-210 as a governance framework for lifecycle evaluation and readjustment (Deutsches Institut für Normung (DIN), 2020b, 2020c). This work creates a citable, auditable basis for aligning the design, evaluation, and institutional allocation of responsibility for digital examinations in a way that is both empirically more precise and normatively more robust.

Problem Exploration: Heterogeneity of Findings and Plurality of Mechanisms

The empirical findings on computer-based testing (CBT) versus paper-and-pencil testing (PPT) are inconsistent in the tertiary context, which should be understood less as a weakness of the research and more as an indication of the context sensitivity of the phenomenon (Al-Fraihat *et al.,* 2020; Eleanor *et al.,* 2014; Erzurumlu, 2023; Ita *et al.,* 2014; Jones *et al.,* 2022; Khoshsima *et al.,* 2017; Oeding, Gunn, & Jones, 2024; Ravesloot *et al.,* 2026; Wagner *et al.,* 2022). For example, a field experimental crossover approach in a university course shows significantly higher performance in computer mode; at the same time, it is plausible that typing allows for longer, more easily editable, and thus more elaborate answers—resulting in an increase in performance that does not necessarily mean more competence in the form of lateral acquisition; it reflects, however, an increase in production and revision capacity under time pressure. (Ceka & O'Geen, 2019).

In contrast, a study conducted in a first-semester engineering setting found no significant differences in performance, with students showing a marked preference for paper—a finding that highlights the difference between outcome equality and experience/stress difference, thus pointing to latent dimensions of fairness

and acceptance (Ita *et al.,* 2014).

Added to this is the integrity and monitoring dimension, which is decisive in practice: remote proctoring can influence performance outcomes (e.g., course grades) and acts as a separate, potentially confounding pathway (monitoring regime → stress/behavioral adjustment → performance) (Andreou *et al.,* 2021; Davis *et al.,* 2016; Hall *et al.,* 2022; Meulmeester *et al.,* 2021; Milone *et al.,* 2017; Oeding, Gunn, & Seitz, 2024; van Halem *et al.,* 2021).

Therefore, it is evident that performance differences can result from a combination of mechanisms (production speed, editability, navigation load, error tolerance, monitoring, and context disruptions) that cannot be reduced to the modality of keyboard vs. pen.

Normative Frame of Reference: ISO 9241 as a grammar of Validity and Fairness

This is precisely where this article comes in: it shifts the explanatory focus from mode to usability and interaction principles within the increasingly digitized environment of tertiary education. DIN EN ISO 9241-11 defines usability as the extent to which specific users can effectively, efficiently, and satisfactorily achieve specific goals in a specific context of use, i.e., with regard to self-conceptual representation, perception, and reception (Deutsches Institut für Normung (DIN), 2018a; Geis & Tesch, 2023; Vindigni, 2026c). However, this concept makes usability a validity condition in testing contexts: if interaction barriers impair goal achievement (response production), they change what the test actually measures (Deutsches Institut für Normung (DIN), 2018a; Vindigni, 2025c, 2026b, 2026c). In the sense of plausibility and bridging arguments (without primary data status), the contextual evidence axis is operationalized using previously published studies on digitally mediated tertiary usage constellations; in particular, findings on cognitive control/load and normatively framed usability in video-based teaching settings as well as on governance and exclusion mechanisms under accessibility regimes and the need for analog fallback architectures are used (Vindigni, 2025c, 2026b, 2026d).

In terms of content, DIN EN ISO 9241-110 specifies the above through interaction principles (including task suitability, self-descriptiveness, conformity with expectations, error tolerance, controllability, customizability, and learning support for the purpose of user-centered and permeable user engagement) (Deutsches Institut für Normung (DIN), 2020b; Vindigni, 2026c). These principles can be operationalized for digital exams and translated into measurable fairness and outcome risks (e.g., scrolling/navigation burden as a violation of task appropriateness; lack of error tolerance as a risk of systematic performance losses for certain groups) (Deutsches Institut für Normung (DIN), 2020b). In this regard, DIN EN ISO 9241-210 ultimately defines digital testing as the outcome of a human-centered design process that connects the context of use, requirements, and evaluation. This process, especially for risky systems, necessitates iterative evaluations that correspond to the level of risk (Deutsches Institut für Normung (DIN), 2020c; Vindigni, 2026c): ex ante (conceptual), interim (processual), and results-oriented (ex post). In tests, the risk is not only functional (system failure) but also normative (misclassification of performance; exclusion) (Deutsches Institut für Normung (DIN), 2020c; Vindigni, 2025f, 2026c).

Research gap: From Comparative Studies to Normative Causal Construction

Although comparative studies (CBT vs. PPT) and reviews on writing mode effects provide valuable evidence, a gap remains: Dominant research often treats mode effects heuristically (as a medium effect) and only rarely as a normatively explicable, socio-technical interaction system that can (a) threaten validity, (b) asymmetrically influence fairness, and (c) shift outcomes via confounding paths (Bashir *et al.,* 2021; Browning *et al.,* 2021; Elmer *et al.,* 2020; Johnson *et al.,* 2020; Khoshsima *et al.,* 2017; Wagner *et al.,* 2022). This paper addresses this gap not with another isolated usability study, but with a triangulation of (i) comparative studies in a tertiary context, (ii) integrity/proctoring research, and (iii) an ISO 9241-based reconstruction of the effect paths (Mogey *et al.* 2010; Ita *et al.* 2014; Deutsches Institut für Normung (DIN) 2018b; Ceka and O'Geen 2019; Deutsches Institut für Normung (DIN) 2020c; DIN Media GmbH 2020b; Van der Weel and Van der Meer 2023; Lestari 2024).

Research Objectives, Research Questions, and Working Hypotheses

Based on established research methodology systematization, it makes perfect sense in the context of this exploratory study to distinguish between primary and secondary objectives and—despite the lack of a separate controlled study—to formulate working hypotheses as normatively justified, empirically verifiable derivations (Arksey & O'Malley, 2005; Daudt *et al.,* 2013; Komljenovic & Williamson, 2024; Snyder, 2019; Torraco, 2005; Wacker, 1998; Whittemore & Knafl, 2005).

In the first instance, the aim is to develop an ISO 9241-based framework model that reconstructs performance outcomes in digital exams as a function of a socio-technical examination system, thereby making validity and fairness risks systematically explainable and assessable (Deutsches Institut für Normung (DIN) 2018b; Deutsches Institut für Normung (DIN) 2020c; DIN Media GmbH 2020b).

Regarding the secondary research objectives, it is necessary to determine the extent to which (1) the synthesis and structuring of heterogeneous findings on CBT vs. PPT in tertiary settings, including extended response formats, can be reconstructed along the ISO categories of context of use, goals, system and interaction quality, and error consequences (Ceka & O'Geen, 2019; Ita *et al.,* 2014; Lestari, 2024). Furthermore, (2) proctoring and integrity must be explained as independent confounding factors

within digital examination arrangements, since they can materially shape assessment outcomes by influencing student behavior, stress conditions, and the overall testing environment (Oeding, Gunn, & Jones, 2024; Oeding, Gunn, & Seitz, 2024). Consequently, (3) the derivation of hypotheses, minimal operationalizations for future studies, and governance- and design-related test criteria for universities must be made explicit (Deutsches Institut für Normung (DIN) 2018b; Deutsches Institut für Normung (DIN) 2020c; DIN Media GmbH 2020b).

Research Questions

Against the backdrop of the normatively guided model and evidence synthesis, the following research questions are formulated not as a linear chain but as complementary guiding questions that together structure the reconstruction of the impact pathways of digital examinations as a socio-technical interaction system and their translation into ISO 9241-based validity and fairness criteria.

Q1: Under what conditions of use can performance differences between handwritten and computer-based exams in tertiary education be interpreted as mode effects in the narrow sense—and when be classified as effects of the socio-technical interaction system?

Q2: Which ISO 9241-110 principles can be used to translate typical CBT design decisions (navigation, input, feedback, and error management) into specific validity and fairness risks?

Q3: What role does remote proctoring play as a monitoring regime in the reconstruction of outcome paths (e.g., via stress/behavioral adaptation, contextual disruptions, technical frictions) and thus as a potential confounder in CBT comparison studies?

Q4: What minimum criteria (design, evaluation, governance) are normatively required for digital exams to be considered sufficiently usable and thus valid in accordance with ISO 9241-11/-210?

Working Hypotheses

Since the contribution establishes an intersubjectively comprehensible rigor via a determinant-indicator-premise architecture in a deductive approach, the derivation of working hypotheses as a translation between normative quality grammar and empirical testing and measurement logic is methodologically justified (Arksey & O'Malley, 2005; Daudt *et al.,* 2013; Komljenovic & Williamson, 2024; Snyder, 2019; Torraco, 2005; Wacker, 1998; Whittemore & Knafl, 2005). Working hypotheses are thus identified as mechanism-based, potentially falsifiable propositions that complement causal-confirmatory test logic and explain the assumptions of effect reconstructed in the synthesis so they can be operationalized in a context-sensitive manner in future CBT studies, tested against boundary conditions, and differentiated with regard to confounding structures (Arksey & O'Malley, 2005; Daudt *et al.,* 2013; Komljenovic & Williamson, 2024; Snyder, 2019; Torraco, 2005; Wacker, 1998; Whittemore & Knafl, 2005).

H1 (direct interaction load): The higher the navigational and editorial interaction load (e.g., scrolling, context switching, lack of overview), the greater the decline in the efficiency of goal achievement and the more likely performance outcomes are to shift independently of the intended construct.

H2 (indirect, error tolerance): Low error tolerance (e.g., risky submit logic, unstable input fields, lack of recoverability) creates asymmetric outcome risks that manifest as fairness problems (user-specific vulnerability).

H3 (direct, production capacity): In extended response formats, the higher production and revision capacity of typing acts as a mechanism that can elaborate responses; observed performance increases in CBT mode are therefore not automatically increases in competence but may reflect response production advantages.

H4 (indirect, acceptance/comfort): Even in the absence of score differences, differences in preference and comfort may indicate latent differences in stress that influence long-term performance outcomes (e.g., via anxiety, concentration stability).

H5 (confounding, monitoring): Proctoring changes the context of use and can act as an independent path to performance outcomes; CBT comparisons without control of the monitoring regime risk spurious explanations.

H6 (processual, governance): The less mature the human-centered design and evaluation process (ISO 9241-210), the higher the probability that digital exams will accumulate undetected interaction problems that only become apparent in outcome artifacts or exclusion.

Contribution, Scope, and Structure of the Article

This article is conceptual and synthetic in nature and aims to provide a normatively controlled reconstruction of the relevant impact pathways; it does not include a separate controlled exam usability study. The epistemic achievement therefore lies in (a) the ISO 9241-based explanation of what is considered a system requirement in digital exams, (b) the integrative rereading of existing findings along these requirements, and (c) the derivation of testable hypotheses and normative test criteria for design and governance (Deutsches Institut für Normung (DIN) 2018b; Deutsches Institut für Normung (DIN) 2020c; DIN Media GmbH 2020b; Lestari 2024; Vindigni 2025b).

In this respect, the contribution does not present primary experimental data from a controlled usability study, a meta-analysis in the strict statistical sense, or an independent causal reevaluation of reported effect sizes. Rather, its research focus lies in the normatively controlled reconstruction of those pathways of influence, relationships between criteria, and contextual conditions under which digital exams, as socio-technical interaction systems, shape the conditions of valid and fair performance representation. In this sense, the paper is conceived as a conceptual-synthetic integration effort that does not bring together disparate findings in the sense of a leveling aggregation; it instead systematizes

this heterogeneity along explicit normative reference categories and translates it into a context that can be evaluated in terms of examination didactics and governance.

MATERIALS AND METHODS

Consequently, this exploratory study is designed as a conceptual-synthetic research contribution with an explicitly normative frame of reference. The aim is not to re-evaluate the effect strength of mode effects in a controlled manner, but rather to provide a normative explanation and integrative reconstruction of the pathways through which digital exams, as socio-technical interaction systems, shape the conditions of valid and fair assessment.

The mode of knowledge can be specified as evidence triangulation: (a) normative axis (ISO 9241 conceptualization of usability, interaction principles, and human-centered design process), (b) empirical axis (findings from comparability studies of CBT vs. paper/handwriting, proctoring/integrity research, and relevant cognitive psychological mechanisms), and (c) contextual axis (connectivity to tertiary examination realities and own, already published digitization findings as plausibility and bridging arguments, without reifying these as primary data sets).

The normative axis is defined by DIN EN ISO 9241-11 (usability as effectiveness/efficiency/satisfaction in the context of use), DIN EN ISO 9241-110 (interaction principles as operationalizable design and evaluation criteria), and DIN EN ISO 9241-210 (human-centered design process including risk-adequate summative evaluation) (Deutsches Institut für Normung (DIN) 2018b; Deutsches Institut für Normung (DIN) 2020c; DIN Media GmbH 2020b).

Moreover, this establishes the central premise that usability in the context of testing should not be treated as a comfort variable but as a condition of validity, insofar as interaction costs, error consequences, and context disruptions can systematically overlap with the intended construct measurement (Deutsches Institut für Normung (DIN), 2018a).

Data Basis: Document Corpus and Inclusion Logic

Accordingly, the material basis to be listed here comprises (i) the aforementioned standard texts of the ISO 9241 series (German versions) as a normative reference and (ii) a thematically focused set of research papers on computer-based vs. paper-based exam formats in higher education, supplemented by proctoring/integrity findings and synthesis-oriented work on writing mode mechanisms (typing vs. handwriting) in extended response formats.

To ensure the reconstruction of mechanisms and confounding factors, different types of evidence are specifically combined. These include (i) field experimental crossover designs, which enable identical or equivalent exam tasks to be completed in a computer-based mode and in a handwritten mode in a university setting, thus contrasting mode conditions while keeping task requirements as constant as possible; (ii) large-scale course studies from the first semester (e.g., engineering sciences), which, in addition to performance outcomes, also map preference, acceptance, and usage experience dimensions, thus revealing the difference between score equality and interaction/stress differences; (iii) synthesis-oriented work on writing mode effects in order to systematically explain the cognitive and text production-related mechanisms (planning, formulation, revision, degree of elaboration) that are often implicit in examination studies; and (iv) quantitative proctoring studies that examine the influence of monitoring regimes on performance indicators (e.g., course grades), thereby addressing proctoring as an independent, potentially outcome-relevant pathway in digital examination settings (Wang *et al.* 2008; Mogey *et al.* 2010; Eleanor *et al.* 2014; Ita *et al.* 2014; Al-Qdah and Ababneh 2017; Ceka and O'Geen 2019; Al-Fraihat *et al.* 2020; Dendir and Maxwell 2020; Andreou *et al.* 2021; Lestari 2024; J. Oeding *et al.* 2024; Oeding *et al.* 2024; Weirich *et al.* 2024; Ravesloot *et al.* 2026).

Accordingly, the inclusion logic does not follow a meta-analytic aggregation goal, nor does it follow a mechanism-oriented selection rationale: Studies are included that (a) relate to examination/assessment contexts (preferably tertiary), (b) report mode/context variations (CBT/paper/handwriting; proctoring regime), (c) provide sufficient information on task/item format, context of use, and implementation, or (d) explain mechanisms (e.g., production/revision dynamics, stress/strain, context disruptions) that are compatible with ISO 9241-based effect path reconstruction. The deliberate inclusion of different publication types is methodologically justified by the fact that mode effect research (especially in the high-risk area) is sometimes reported outside of classic journal logic and that mechanism reconstruction is not exclusively linked to peer-reviewed RCT designs (Browning *et al.,* 2021; Johnson *et al.,* 2020; Jones *et al.,* 2022; Lestari, 2024; Wagner *et al.,* 2022; Yanfei, 2021).

Analysis Strategy: ISO Mapping, Path Reconstruction, and Synthesis

The analysis strategy used here does not implement triangulation as a multiple method per se, i.e., in the sense of parallel primary surveys; it follows, however, a structured assignment of heterogeneous evidence along a normative reference grid (ISO mapping) and a reconstructive modeling of plausible impact pathways between system characteristics, usage context, interaction actions, and performance outcomes.

a) ISO 9241-11 mapping (context of use → usability components):

Findings are first organized according to the 9241-11 categories of users, goals/tasks, resources/environment, and context conditions and then related to the usability components of effectiveness, efficiency, and satisfaction in order to reveal systematic interaction costs as a potential

validity condition (Deutsches Institut für Normung (DIN), 2018a; Vindigni, 2026c).

b) ISO 9241-110 mapping (principles → risks/operationalizations):

The findings are then related to the interaction principles of 9241-110 (including task appropriateness, self-descriptiveness, conformity to expectations, error tolerance, and controllability). This enables design decisions for digital exams to be translated into measurable fairness and outcome risks (e.g., lack of error tolerance → irreversible loss of answers; navigation burden → increased processing time/probability of errors; controllability → loss of control in restrictive proctoring setups) (Deutsches Institut für Normung (DIN), 2020b; Geis & Tesch, 2023; Vindigni, 2026c).

c) ISO 9241-210 process perspective (HCD → governance/evaluation logic):

To do this, the framework model is supplemented by the procedural perspective of 9241-210, especially where digital examinations as high-risk applications require a risk-adequate summative evaluation and the assignment of responsibility on the part of the organization (Deutsches Institut für Normung (DIN), 2020c; Geis & Tesch, 2023; Vindigni, 2026c). The standard provides explicit process requirements for this (including the scope of evaluation depending on the risk of non-compliance with requirements, inspection and user-based evaluation, and long-term observation), which are operationalized in the article as governance and minimum requirement logic (DIN Media GmbH 2020b).

d) Impact pathway reconstruction and confounding diagnostics:

Finally, recurring pathway patterns can be reconstructed on the basis of ISO mapping, for example:

• Input/editable → Response length/elaboration → Score (especially for extended responses) (Al-Fraihat *et al.,* 2020; Ceka & O'Geen, 2019).

• Proctoring regime → Stress/behavioral adjustment/context disruption → Performance (Andreou *et al.,* 2021; Davis *et al.,* 2016; Dendir & Maxwell, 2020; Gibson & Kuhn, 2024; Hall *et al.,* 2022; Malik *et al.,* 2023; Meulmeester *et al.,* 2021; Milone *et al.,* 2017; Murphy *et al.,* 2022; Oeding, Gunn, & Seitz, 2024; Ravesloot *et al.,* 2026; van Halem *et al.,* 2021; Woldeab & Brothen, 2024).

• Navigation/display load → Efficiency/error costs → Outcome (also in the case of score equality as a signal of fairness/acceptance).

Nevertheless, it should be noted that these paths are not claimed to be causal effect estimates but are formulated as analytical propositions that essentially (a) highlight plausible confounders in CBT vs. paper comparisons and (b) guide more selective future operationalizations.

Regarding the significance of the present reconstruction, it must be noted that it is limited insofar as the underlying findings constitute a heterogeneous evidence base; the systematization undertaken here is not designed as a meta-analysis in the strict statistical sense but instead as a normatively controlled synthesis of various empirical and conceptual contributions. In this respect, the research contribution, therefore, does not claim to provide a direct causal quantification of individual influencing factors but rather a theory-guided explanation of possible causal relationships that can be examined in detail and operationalized within the framework of further empirical research. In particular, it should be noted that the reconstruction is based on published studies and normative reference frameworks and therefore cannot fully capture context-dependent variations in digital examination architectures. Nevertheless, the strength of the chosen approach lies in translating previously fragmented findings into an explicit framework that allows for the systematic identification of validity, fairness, and exclusion risks associated with digital exams and makes them compatible with exam pedagogy.

Transparency and Reporting Standard: PRISMA Principle and Flow Diagram

Initially, it should be noted that, although the exploratory study described here is not strictly designed as a traditional systematic review with a meta-analysis, the PRISMA guidelines are in fact used as a reporting framework to ensure transparency and reproducibility, taking into account research ethics implications and quality criteria, particularly regarding the documentation of

Table 1: PRISMA Flow Data (for Diagram).

Stage	Description	n
Identification	Records identified through databases/sources: Scopus = 1200; Web of Science Core Collection = 950; ERIC (Education Resources Information Center) = 600; PsycINFO = 500; PubMed/MEDLINE = 400; CINAHL (EBSCO) = 250; IEEE Xplore = 350; ACM Digital Library = 300; INSPEC = 300; Education Source (EBSCO) = 350; EBSCO Academic Search Complete = 350; SocINDEX (EBSCO) = 200; ProQuest Central = 200; ProQuest Dissertations & Theses = 210; ScienceDirect (Elsevier) = 500; SpringerLink = 400; Wiley Online Library = 300; Taylor & Francis Online = 350; SAGE Journals = 260; Emerald Insight = 170; JSTOR = 140; arXiv (preprints) = 200; SSRN (working papers) = 220; DOAJ (Directory of Open Access Journals) = 120; BASE (Bielefeld Academic Search Engine) = 220; OpenAIRE = 140; CORE = 110; Google Scholar (grey literature, reports, preprints) = 326; W3C/WAI resources (WCAG, techniques, supporting docs) = 60; ETSI (EN 301 549 and related deliverables) = 40; EUR-Lex (EAA/BFSG-related legal texts and amendments) = 30	9746

Duplicates removed	Records automatically and manually deduplicated across all databases/sources	2380
Records screened (titles/abstracts)	After deduplication; title and abstract screening for relevance to tertiary education exams (CBT vs. paper/handwriting), mode effects, proctoring/integrity, usability/HCI/ergonomics, accessibility (WCAG/BITV/EN 301 549), and assessment validity/fairness	7366
Records excluded	Excluded due to out-of-scope topic, non-tertiary-only focus without transferability, purely technical testing without assessment context, insufficient conceptual/methodological substance, or lacking relevance to at least one analytical complex (assessment outcomes; socio-technical interaction conditions; governance/integrity/accessibility)	6550
Reports sought for retrieval	Full texts sought for retrieval after title/abstract screening	816
Reports not retrieved	Full texts not retrieved / not accessible (e.g., paywalled without institutional access, incomplete records)	54
Reports assessed for eligibility	Full-text assessment in relation to (a) CBT vs. paper/handwriting comparability and/or writing-mode mechanisms, (b) proctoring/integrity as context intervention, and (c) socio-technical interaction conditions (context of use, system quality, error consequences, accessibility/compliance)	762
Full-text reports excluded	Exclusion reasons at full text: out of scope (n = 380); non-tertiary-only without justified transfer (n = 140); insufficient method/reporting transparency (n = 105); non-comparative/editorial or purely opinion-based formats (n = 95)	720
Studies included in qualitative synthesis	Methodologically and substantively relevant studies/documents for synthesis (comparability evidence; proctoring/integrity; cognitive mechanisms; socio-technical interaction and accessibility conditions)	42
Studies included in conceptual synthesis	Subset integrated into the final ISO-9241-based framework model, mapping matrix, and proposition set	24

Source: Author (2026). Proposed PRISMA flow data for diagram (counts for planning and reporting structure; search window: 2008–Feb 2026)

identification, screening, eligibility criteria, and inclusion (PRISMA Executive, 2020). To this purpose, a PRISMA flow diagram in accordance with PRISMA 2020 is provided, which transparently illustrates the source flows (databases/registers, manual search, reference lists, and exclusions, including justification categories) (Haddaway *et al.,* 2022; Page *et al.,* 2021; PRISMA Executive, 2020). For transparency, the PRISMA-oriented source flows

Table 2: PRISMA Records Identified Breakdown

Source	**n**
Scopus	1200
Web of Science Core Collection	950
ERIC (Education Resources Information Center)	600
PsycINFO	500
PubMed/MEDLINE	400
CINAHL (EBSCO)	250
IEEE Xplore	350
ACM Digital Library	300
INSPEC	300
Education Source (EBSCO)	350
EBSCO Academic Search Complete	350
SocINDEX (EBSCO)	200
ProQuest Central	200
ProQuest Dissertations & Theses	210
ScienceDirect (Elsevier)	500

SpringerLink	400
Wiley Online Library	300
Taylor & Francis Online	350
SAGE Journals	260
Emerald Insight	170
JSTOR	140
arXiv (preprints)	200
SSRN (working papers)	220
DOAJ (Directory of Open Access Journals)	120
BASE (Bielefeld Academic Search Engine)	220
OpenAIRE	140
CORE	110
Google Scholar (grey literature, reports, preprints)	326
W3C/WAI resources (WCAG, techniques, supporting docs)	60
ETSI (EN 301 549 and related deliverables)	40
EUR-Lex (EAA/BFSG-related legal texts and amendments)	30
Total	9746

Source: Author (2026)

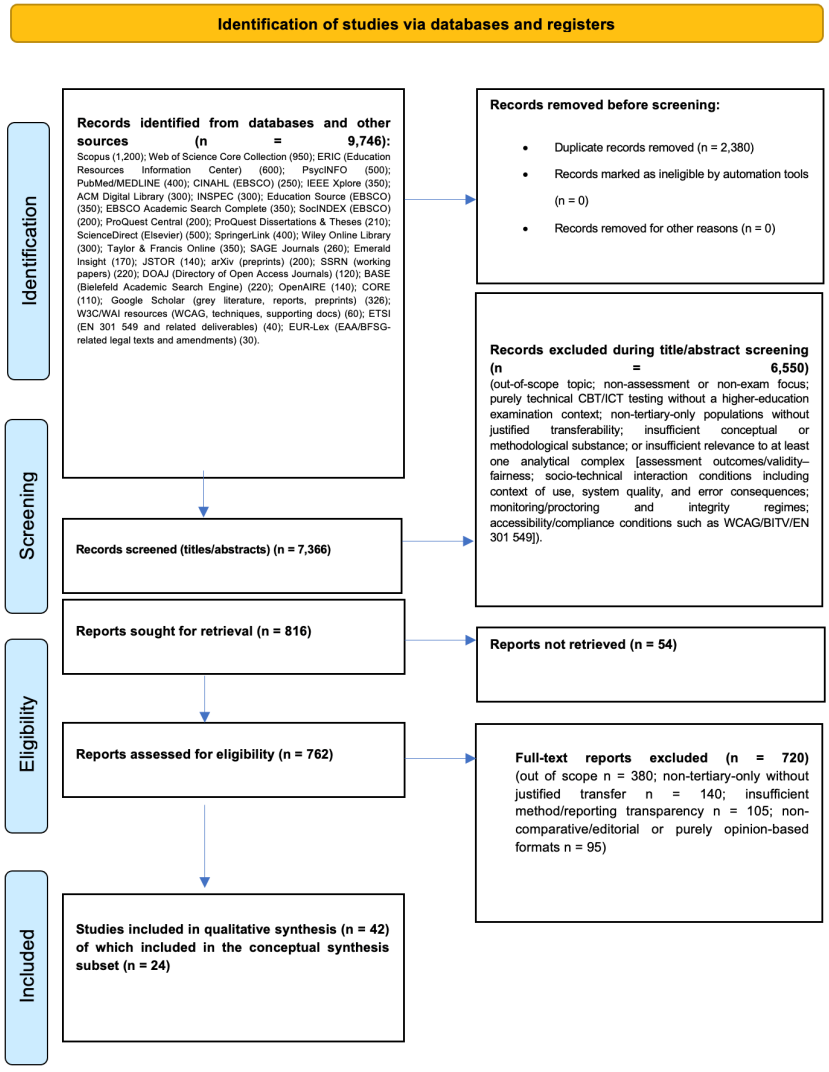

Figure 1: PRISMA flow diagram (adapted from PRISMA 2020) for documenting the identification, screening, eligibility assessment, and inclusion of studies included in the evidence triangulation.
Source: Author (2026)

are summarized in Table 1 and Table 2 and visualized in Figure 1.

Method results: model, matrix, and agenda

The method leads to three closely linked results:

a) An ISO 9241-based framework model (conceptual figure) that reconstructs performance outcomes as a function of the socio-technical testing system;

b) A mapping matrix that assigns the findings to the ISO categories (9241-11/110/210) and the identified path types (Deutsches Institut für Normung (DIN) 2018b; Deutsches Institut für Normung (DIN) 2020c; DIN Media GmbH 2020b); and

c) A research agenda in the form of testable propositions/working hypotheses, including minimal operationalizations, that can be used in future studies to identify confounding factors in CBT settings with greater precision.

Results

Findings and Types of Evidence

The literature reviewed does not form a homogeneous body of evidence; it instead constitutes a bundle of partially incommensurable research logics that require a specific systematic integration: (a) comparative performance studies on CBT vs. paper-and-pencil or handwritten exams, (b) integrity/proctoring studies focusing on monitoring regimes, opportunities for cheating, and their performance artifacts, and (c) cognitive psychology studies on the mechanisms of handwriting vs. keyboard input (including memory, planning/formulation processes, and stress) (Al-Qdah & Ababneh, 2017; Ceka & O'Geen, 2019; Eleanor *et al.,* 2014; Erzurumlu, 2023; Ita *et al.,* 2014; Lestari, 2024; Mogey *et al.,* 2010; Van der Weel & Van der Meer, 2023; Wang *et al.,* 2008; Weirich *et al.,* 2024). This heterogeneity is not a shortcoming for the present work, rather, it constitutes the methodological starting point for triangulation. The findings are not averaged additively; instead, they are reconstructed along plausible pathways of effect that are normatively based on the ISO 9241 terms of usability in the context of use (DIN EN ISO 9241-11) as well as the dialogue principles (DIN EN ISO 9241-110) and the human-centered design/evaluation process (DIN EN ISO 9241-210) (Al-Fraihat *et al.,* 2020; Al-Qdah & Ababneh, 2017; Andreou *et al.,* 2021; Ceka & O'Geen, 2019; Deutsches Institut für Normung (DIN), 2018a, 2020b, 2020c; Geis & Tesch, 2023; Hall *et al.,* 2022; Ita *et al.,* 2014; Lestari, 2024; Vindigni, 2026c; Weirich *et al.,* 2024).

Performance Differences (mode effects) as an Observable but Insufficiently Explained Phenomenon

Despite all context-specific fluctuations, the comparative studies reveal a recurring pattern: performance differences between computer-based and paper/handwritten exam formats are not consistently explained by the input modality alone but occur depending on the task format, answer length, time regime, system stability, writing/editing requirements, and exam environment (Al-Qdah & Ababneh, 2017; Andreou *et al.,* 2021; Ceka & O'Geen, 2019; Cramp *et al.,* 2019; Eleanor *et al.,* 2014; Erzurumlu, 2023; Ita *et al.,* 2014; Lestari, 2024; Milone *et al.,* 2017; Mogey *et al.,* 2010; Newton, 2023; Noorbehbahani *et al.,* 2022; Van der Weel & Van der Meer, 2023; van Halem *et al.,* 2021; Wang *et al.,* 2008; Woldeab & Brothen, 2024). Accordingly, the mode effect often remains a descriptive residue: it is reported but not broken down into the interaction and context conditions that shift it toward validity, fairness, and outcome (Baragash & Al-Samarraie, 2018; Deutsches Institut für Normung (DIN), 2020b, 2020c; Johnson *et al.,* 2020; Yanfei, 2021).

A meta-analytical perspective from comparability studies (albeit partly outside the tertiary sector) is instructive in this regard because it methodologically reveals that mode is often operationalized as a test administration configuration (device, interface, navigation, timing), while usability as a validity condition is rarely explicitly modeled (Malik *et al.,* 2023; Milone *et al.,* 2017; Oeding, Gunn, & Jones, 2024; Oeding, Gunn, & Seitz, 2024; Woldeab & Brothen, 2024). The consequence that can be derived from this reasoning is that where the test environment is variable, an apparent mode effect may in fact be a bundle of confounding factors (Malik *et al.,* 2023; Milone *et al.,* 2017; Oeding, Gunn, & Jones, 2024; Oeding, Gunn, & Seitz, 2024; Wang *et al.,* 2008; Woldeab & Brothen, 2024).

Mediation Paths: From Input Handling to Outcome Artifacts

In terms of facts, the research methodological triangulation suggests that performance differences—if they occur de facto—are often mediated by interaction effects. Four path clusters can be reconstructed in the studies:

a)Interaction load and attention binding

Digital exams generate additional side tasks (interface navigation, cursor/window control, formatting decisions, and error handling). In ISO 9241-11 terminology, this does not primarily concern the whether of goal achievement, but rather the efficiency and satisfaction under specific usage context conditions (Al-Fraihat *et al.,* 2020; Ceka & O'Geen, 2019; Deutsches Institut für Normung (DIN), 2018a; Hall *et al.,* 2022).

b) Error consequences and error tolerance

In CBT settings, mistakes (e.g., accidental overwriting, loss of text, submission of incorrect versions, session timeouts) can have different consequences than in paper format. This shifts the outcome variance from content competence to system-induced error risks. ISO 9241-110 provides the normative translation of such risks into testable principles (including error tolerance, controllability, and conformity with expectations). (Al-Fraihat *et al.,* 2020; Al-Qdah & Ababneh, 2017; Andreou *et al.,* 2021; Ceka & O'Geen, 2019; Dendir & Maxwell, 2020; Deutsches Institut für Normung (DIN), 2020b; Eleanor *et al.,* 2014; Ita *et al.,* 2014; Ravesloot *et al.,* 2026; Vindigni, 2025c; Wang *et al.,* 2008; Weirich *et al.,* 2024).

c) Context of use and situational stress

The context of use (end device quality, input device, network, room/supervision situation, physical ergonomics) also acts as a conditioning factor for interaction. In a socio-technical interpretation, mode is thus only the surface of a context that influences the distribution of cognitive resources during processing (and thus affects validity and fairness) (Al-Fraihat *et al.,* 2020; Browning *et al.,* 2021; Ceka & O'Geen, 2019; Elmer *et al.,* 2020; Hall *et al.,* 2022; Lestari, 2024; Mogey *et al.,* 2010; Wagner *et al.,* 2022; Wang *et al.,* 2008).

d) Writing process mechanisms (handwriting vs. typing)

In this context, the relevant literature points out that handwriting and typing can support different microprocesses of text production, planning, and control; however, these mechanisms are often not surveyed together with interface/context variables in examination studies. This creates a central research gap: it also explains why the cognitive mechanisms are discussed in current research without being systematically linked to ISO 9241-compliant interaction conditions of the specific exam software and thus remain deficient (Al-Qdah & Ababneh, 2017; Andreou *et al.,* 2021; Ceka & O'Geen, 2019; Cramp *et al.,* 2019; Deutsches Institut für Normung (DIN), 2020b; Erzurumlu, 2023; Lestari, 2024; Masterman & Fresen, n.d.; Milone *et al.,* 2017; Mogey *et al.,* 2010; Newton, 2023; Noorbehbahani *et al.,* 2022; van Halem *et al.,* 2021; Woldeab & Brothen, 2024).

Remote Proctoring as a Performance and Fairness Intervention

From a consensus theory perspective, the reviewed investigator-triangulated research exhibits varying results: in certain scenarios, proctoring leads to a decline in performance, while in others, no significant differences are observed; additionally, stress and acceptance dimensions such as privacy, test anxiety, and emotional distress become apparent (Andreou *et al.,* 2021; Davis *et al.,* 2016; Dendir & Maxwell, 2020; Gibson & Kuhn, 2024; Hall *et al.,* 2022; Meulmeester *et al.,* 2021; Milone *et al.,* 2017; Murphy *et al.,* 2022; Oeding, Gunn, & Jones, 2024; Oeding, Gunn, & Seitz, 2024; Ravesloot *et al.,* 2026; van Halem *et al.,* 2021; Woldeab & Brothen, 2024).

This finding is particularly central to the framework model because monitoring not only addresses integrity but also expands the socio-technical system to include a controlling interaction actor that ties up resources and can induce stress (Andreou *et al.,* 2021; Milone *et al.,* 2017; Oeding, Gunn, & Jones, 2024; Oeding, Gunn, & Seitz, 2024; Woldeab & Brothen, 2024).

A large-scale study in the graduate sector (N≈426) reports a statistically significant influence of the introduction of proctoring software on course grades (when controlling for GPA, among other factors). This means that proctoring must be taken seriously as a potential outcome driver—regardless of whether the mechanism is mediated by actual fraud reduction or by changed stress/processing conditions (Oeding, Gunn, & Jones, 2024; Oeding, Gunn, & Seitz, 2024).

At the same time, the question of the validity of unproctored online summative exams is being hotly debated: if exam integrity cannot be guaranteed, the interpretation of performance outcomes as a performance indicator becomes fundamentally fragile. Specifically, this means that for the framework model to be explored here, validity must be understood not only in terms of test theory but also as a system property per se (including governance) (Newton, 2023).

ISO 9241-based reconstruction: From findings to normative test criteria

As a result, empirical heterogeneity can be converted into a normatively testable structure:

a) ISO 9241-11 (usability in the context of use) acts as a validity condition: If effectiveness/efficiency/satisfaction varies systematically depending on the context (device, network, interface, proctoring), then the measurement of the intended competence may also vary (Ceka & O'Geen, 2019; Deutsches Institut für Normung (DIN), 2018a; Geis & Tesch, 2023; Vindigni, 2026c).

b) ISO 9241-110 provides the logic for translating interaction deficiencies into fairness and outcome risks: Error tolerance, conformity to expectations, and controllability become operationalizable design/evaluation criteria, the violation of which makes systematic disadvantages plausible (e.g., for certain writing/input routines, assistive technology use, and stress sensitivity) (Deutsches Institut für Normung (DIN), 2020b; Eleanor *et al.,* 2014; Geis & Tesch, 2023; Ita *et al.,* 2014; Vindigni, 2025c, 2026c).

c) ISO 9241-210 establishes the governance requirement to treat digital exams as systems to be designed and evaluated iteratively (understand the context of use, specify requirements, iterate solutions, anchor evaluation)—not as a usability add-on, but as an integral part of exam-related quality assurance (Deutsches Institut für Normung (DIN), 2020c; Geis & Tesch, 2023; Lestari, 2024; Vindigni, 2026c).

In this regard, this allows the central finding to be formulated precisely: performance differences are often less to be interpreted as a modality effect and more as a system effect resulting from the context of use, interaction quality, error consequences, monitoring regime, and (implicit) accessibility conditions; it is precisely these system effects that can be explicitly defined and evaluated in normative terms.

Presentation of Results and Documentation Artifacts

Rather than resulting in an additive summary of heterogeneous individual results, the triangulation presented here leads to a structured representation of results that translates the reconstruction of impact pathways and confounding factors into a verifiable and governable form. In this sense, the central results are condensed into two complementary presentation formats that make both the normative logic (ISO 9241) and the empirical connectivity of the synthesis transparent.

First, the ISO 9241-based framework model is presented

as an integrative system representation (see Figure 2). It visualizes digital exams as socio-technical examination arrangements in which usage context conditions (DIN EN ISO 9241-11), interaction principles (DIN EN ISO 9241-110), and process and governance requirements (DIN EN ISO 9241-210) function as structuring levels, the intertwining of which can give rise to validity, fairness, and outcome risks (Deutsches Institut für Normung (DIN), 2018a, 2020b, 2020c). This does not level out the heterogeneity of the findings; it instead explains this variance as the constellation-dependent activation of specific path types.

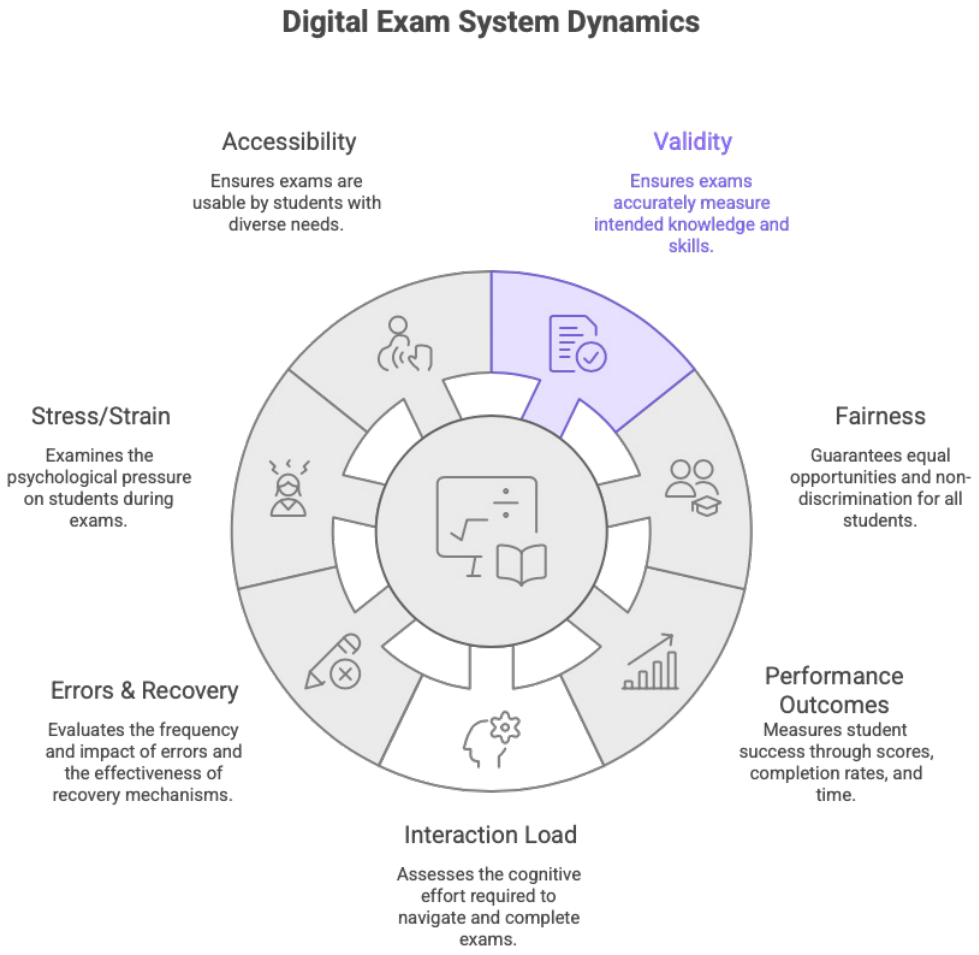

Figure 2: DIN EN ISO 9241–anchored framework model of the socio-technical examination system in tertiary education
Source: Author (2026)

Second, the translation performance of the synthesis is documented in a mapping and path matrix (see Supplement C for the full version), which assigns reconstructed impact paths (e.g., interaction load, error tolerance/error consequences, proctoring as a contextual intervention, accessibility, and resource conditions) to the respective ISO anchors as well as to derived risk and operationalization classes; a condensed representation is provided in the main text where appropriate.

The content is deliberately designed as an auditable and evaluable summary: it combines (i) normatively specified quality dimensions, (ii) typical failure modes of digital examination environments, and (iii) testable indicators that can be used in future studies to identify confounding factors in CBT settings with greater precision.

Also, for methodological transparency, the composition of the study corpus used is disclosed in accordance with the PRISMA principle (see Figure 1, PRISMA flow) (Haddaway *et al.,* 2022; Page *et al.,* 2021; PRISMA Executive, 2020). In addition, the search and screening logic, the extraction scheme, and the complete mapping matrix are documented in the appendices (see Appendices A–C / Supplementary Materials) to ensure the replicability and traceability of the normatively guided evidence exploration.

Following the reconstruction of the effect pathways, the results are condensed into two complementary forms of representation. Supplement C documents (i) the identified effect paths as an assignment of typical triggers in CBT settings, (ii) associated validity and fairness risks, and (iii) the relevant ISO 9241 anchors (DIN EN ISO 9241-11/-110/-210) and exemplary evidence. Figure 2 presents a model of the socio-technical examination system derived from this and consolidates the triangulation by identifying the examination task, examinee, CBT system (UI/technology), proctoring layer, and usage context as an interaction arrangement from which outcome-

relevant risks can arise. To strengthen intersubjective traceability and to mitigate systematic distortions in the synthesis process, the complete classification matrix and the extraction scheme are reported in the appendix as an explicit audit trail, in line with established journal reporting conventions.

Notes

The figure conceptualizes digital exams as interaction arrangements in which performance outcomes emerge from the coupled effects of context of use, system quality, error consequences, monitoring/proctoring regimes, and accessibility conditions rather than from input modality alone. In line with DIN EN ISO 9241-11, usability in the specified context (effectiveness, efficiency, satisfaction) is treated as a constitutive validity condition because interaction barriers can introduce construct-irrelevant variance and distort what is measured. DIN EN ISO 9241-110 operationalizes this translation through dialogue principles (e.g., task suitability, self-descriptiveness, controllability, and error tolerance), enabling exam-relevant risks to be mapped to concrete interface and workflow properties. DIN EN ISO 9241-210 frames these requirements as a human-centered design and evaluation cycle, positioning continuous requirement specification and empirical evaluation as governance mechanisms of the exam service across its life cycle. The model thereby links normative criteria to empirically observable pathways (interaction load, stress/strain, recovery effort, exclusion risk) and to the outcome dimensions of validity, fairness, and performance.

Summary of Results: Key Findings and Core Implications

Overall, the triangulation shows that differences in performance between handwritten and computer-based exams in tertiary education can only be interpreted as pure modality effects in exceptional cases and should instead be reconstructed primarily as interaction effects of socio-technical exam arrangements (Baragash & Al-Samarraie, 2018; Browning *et al.,* 2021; Ceka & O'Geen, 2019; Eleanor *et al.,* 2014; Ita *et al.,* 2014; Jones *et al.,* 2022; Wagner *et al.,* 2022; Wang *et al.,* 2008; Weirich *et al.,* 2024). These interaction effects emerge from the intertwining of usage context (resources, environment, device/network quality), interaction quality (navigation, feedback, controllability), error consequences (irreversibility, recoverability), and—where relevant—monitoring regimes (remote proctoring) and implicit accessibility conditions (Ceka & O'Geen, 2019; Eleanor *et al.,* 2014; Ita *et al.,* 2014; Milone *et al.,* 2017; Oeding, Gunn, & Jones, 2024; Oeding, Gunn, & Seitz, 2024).

This gives ISO 9241 a dual function: DIN EN ISO 9241-11 is to be understood as a validity condition, insofar as effectiveness, efficiency, and satisfaction in the context of use help determine processing performance and thus the construct interpretation of the results (Ceka & O'Geen, 2019; Deutsches Institut für Normung (DIN), 2018a; Geis & Tesch, 2023; Vindigni, 2026c, 2026b).

At the same time, DIN EN ISO 9241-110 provides a principle-based transformation grid that can be used to translate specific design decisions for digital exams (particularly with regard to task appropriateness, error tolerance, conformity with expectations, and controllability) into operationalizable fairness and outcome risks (Andreou *et al.,* 2021; Deutsches Institut für Normung (DIN), 2020b; Eleanor *et al.,* 2014; Geis & Tesch, 2023; Ita *et al.,* 2014; Vindigni, 2025c, 2026b, 2026c).

Finally, DIN EN ISO 9241-210 establishes the governance implication of treating digital exams as systems that should be designed iteratively and evaluated in a manner appropriate to the risks involved, rather than implementing them as a mere administrative variant of traditional exams (Deutsches Institut für Normung (DIN), 2020c; Geis & Tesch, 2023; Lestari, 2024; Vindigni, 2025c, 2026c, 2026b).

From an educational and institutional perspective, this implies that digital exams must not be treated merely as a technical means of delivering traditional exams; instead, they should be regarded as independent, risk-sensitive examination infrastructures whose design quality directly influences the conditions for valid and fair assessment; this is consistent with a broader accessibility-governance perspective in which assessment regimes form part of the socio-technical architecture of democratic and inclusive higher education (Vindigni, 2026d). This gives rise, first, to the necessity of not merely providing digital examination environments in a functional manner but of systematically evaluating them under real-world usage contexts for interaction burdens, consequences of errors, accessibility requirements, and potential risks of exclusion. Second, where such risks cannot be adequately controlled, compensatory or alternative examination arrangements should be understood not as exceptions to a deficient practice but as an expression of responsibility in examination pedagogy and governance.

Thus, the resulting contribution lies in the normatively controlled reconstruction of those path types through which apparent mode effects can in reality represent confounding bundles and in the provision of a citable framework model that translates empirical heterogeneity into a testable structure of validity, fairness, and governance criteria (Aini *et al.,* 2021; Browning *et al.,* 2021; Ceka & O'Geen, 2019; Elmer *et al.,* 2020; Johnson *et al.,* 2020; Jones *et al.,* 2022; Khoshsima *et al.,* 2017; Newton, 2023; Wagner *et al.,* 2022).

Proposition Status Mapping: Evidence Status, Boundary Conditions, and Operationalization Implications

The working hypotheses formulated in the introduction (H1–H6) are not evaluated through confirmatory hypothesis testing. For the purposes of the results synthesis, they are consolidated into six evidence-based propositions (P1–P6), organized by reconstructed

effect-path types rather than by the original hypothesis numbering. This avoids treating the synthesis as a direct hypothesis test and clarifies that the status mapping translates the initial assumptions into a proposition set within the ISO 9241-based framework. The propositions are differentiated by (i) their degree of plausibility, (ii) boundary conditions/moderators, and (iii) operationalization options.

Methodologically, this procedure follows established approaches to integrative and scoping-oriented synthesis in which heterogeneous evidence types are combined to reconstruct mechanisms and to derive testable propositions rather than to estimate causal effects (Arksey & O'Malley, 2005; Daudt *et al.,* 2013; Komljenovic & Williamson, 2024; Snyder, 2019; Torraco, 2005; Whittemore & Knafl, 2005). Accordingly, the present mapping does not claim causal-confirmatory test logic; it explicates a coherent proposition set and its boundary conditions in line with theory-building criteria and conceptual rigor (Wacker, 1998).

Accordingly, a status mapping is carried out that (i) indicates the evidence status of the respective effect assumption across the triangulated evidence types, (ii) explicates central moderators and constraints, and (iii) specifies minimal operationalizations through which future studies can test the propositions with greater precision. In this way, the results logic is consistently presented as a reconstruction of impact pathways in a socio-technical examination system and simultaneously translated into ISO 9241 categories.

(P1) Interaction load as a construct-relevant source of variance—status: largely plausibilized (robust), yet strongly context- and design-dependent.

The evidence indicates that performance differences in CBT settings frequently cannot be isolated as a pure modality effect (keyboard vs. handwriting); they instead emerge from increased interaction load (navigation demands, editor handling, split attention, and workflow frictions). Normatively, this mechanism is directly compatible with DIN EN ISO 9241-11: interaction barriers reduce effectiveness and efficiency and thereby shift the exam situation from measuring competence to partially measuring interaction management. Key boundary conditions include task format (essay vs. short answer vs. MC), UI architecture (editor functions, navigation logic), device heterogeneity/BYOD, network stability, and time regime/timeout design. Minimal operationalization options comprise (i) time spent on navigation/formatting versus content production, (ii) interaction events (e.g., backtracking, cursor jumps, copy-paste restrictions), (iii) short-scale measures of perceived friction/strain, and (iv) abandonment and incompleteness rates.

(P2) Error tolerance and error consequences as outcome risk—status: largely plausibilized (robust), particularly via error consequences rather than error frequency alone.

The results indicate that digital exam environments influence performance not merely through more errors, but through error-consequence regimes, including recovery options, saving reliability (autosave/persistence), undo and confirmation logic, timeout/auto-submit mechanisms, and sanction rules triggered by technical events. In this respect, DIN EN ISO 9241-110 provides the central translation bridge: error tolerance and self-descriptiveness and controllbility function as exam-relevant validity and fairness conditions because they determine whether errors remain construct-irrelevant or are converted into score penalties through recovery friction, time loss, or procedural sanctions. Key boundary conditions include system stability and resilience (crash/disconnect), institutional rules for resumption/restart, support availability, logging practices, and dispute/appeal procedures. Minimal operationalization options comprise (i) error events (system- vs. user-generated), (ii) recovery time, (iii) data loss and submission failures, (iv) number and type of emergency interventions, and (v) sanction triggers (timeouts, lockdown violations, proctoring flags).

(P3) Monitoring/proctoring as an independent pathway—status: largely plausibilized (medium to high), with high confounding and moderation relevance.

The triangulation indicates that proctoring cannot be treated as a neutral integrity layer but must be modeled as a contextual intervention with its own outcome-relevant pathway (stress/strain, inhibition and disruption mechanisms, and false-positive events). From a socio-technical systems perspective, this introduces a double threat to validity and fairness, that means: (i) direct performance effects through monitoring-induced strain and interruptions, and (ii) indirect distortions through altered work behavior (risk avoidance, reduced revision, strategic slowing). Moreover, key boundary conditions include proctoring type (live, recorded, or AI-based), strictness and false-alarm rates, technical lockdown constraints, rule transparency, appeal/review procedures, and cultural/psychological dispositions (e.g., test anxiety). Minimal operationalization options comprise (i) proctoring events/flags and their resolution status, (ii) interruption counts and intervention frequency, (iii) stress/strain indicators (short psychometric scales), (iv) processing speed and revision behavior (edit logs/keystroke dynamics), and (v) grade and completion differences between proctored and non-proctored settings while holding task difficulty constant and controlling for relevant covariates.

(P4) Accessibility conditions as a fairness and exclusion mechanism—status: largely plausibilized (high) and governance-critical.

The synthesis supports the proposition that accessibility should not be treated as a secondary quality feature but as a constitutive condition of fairness and participation in digital exams. Barriers (assistive-technology incompatibilities, insufficient keyboard operability, inadequate contrast and zoom stability, and rigid time-limit regimes) generate systematic risks of exclusion and performance distortion and thus produce construct-

irrelevant variance in validity terms (Vindigni, 2026d). Normatively, this aligns with DIN EN ISO 9241-11 (effectiveness in a specified context of use) and with a governance-oriented interpretation of examination infrastructures: participation must not be made contingent upon implicit device, competence, or assistive-technology prerequisites. Key boundary conditions include institutional implementation of WCAG/BITV/EN 301 549 obligations, quality-assurance routines, procurement and platform selection, formal accommodations procedures, support processes, and right-to-analog fallback options in critical usage contexts. Minimal operationalization options comprise (i) WCAG/EN 301 549 conformance checks using exam-relevant use cases, (ii) assistive-technology compatibility testing, (iii) abandonment and error rates among examinees using assistive technologies, (iv) accommodation processing times and effectiveness indicators, and (v) interaction-friction measures (e.g., time-on-task variance, navigation/recovery events) stratified by accessibility needs.

(P5) Resource and context variance in tertiary education—status: largely plausibilized (medium to high), as a structurally constitutive moderator set of the overall system.

The findings indicate that variance in tertiary usage contexts (BYOD heterogeneity, spatial conditions, network quality, exam location/remote setting, and availability of support) is not mere noise but a structural moderator structure shaping interaction load and outcome distributions. Consequently, score equality can coexist with unevenly distributed interaction burden; fairness problems may then remain invisible in mean-score comparisons but become evident in variance and tail risks, including dropouts, timeouts, and disparities in error and recovery profiles. Key boundary conditions include study type (on-campus vs. distance learning), socioeconomic resources, institutional standardization (managed devices vs. BYOD), support and assistance architectures, and the exam environment. Minimal operationalization options comprise (i) device and network profiles, (ii) context surveys (workplace quality, disruptions), (iii) support-contact rates and escalation events, (iv) distributional analyses (variance and quantile-based comparisons, not only mean values), and (v) disparity analyses stratified by resource strata and accessibility needs.

(P6) Writing-mode mechanisms: text production, revision, and elaboration depth—status: partially plausibilized (medium), highly contingent on task format and editor architecture.

Mechanism-oriented research suggests that handwriting and typing can favor different profiles of planning, formulation, and revision; however, in exam contexts these effects can rarely be isolated because writing mode is typically coupled with UI constraints (editor functionality, navigation burden, copy/paste restrictions) and proctoring conditions. Methodologically, this implies that writing mode is relevant as a sub-path within the socio-technical system but is unlikely to function as a monocausal explanation; its contribution can only be interpreted meaningfully within system reconstruction while controlling for interaction load and context-of-use conditions. Key boundary conditions include essay share and task demands, permitted revision/editing options, keyboard proficiency, time pressure, proctoring restrictions, and language/writing skills. Minimal operationalization options comprise (i) keystroke-based revision metrics (edit events, deletion/insert ratios, and revision bursts); (ii) text quality assessed via analytic rubrics; (iii) processing time, revision rate, and completion indicators (length/completeness); (iv) short measures of strategy and perceived effort (planning/revision); and (v) interaction models capturing moderation effects (writing mode × UI design × proctoring).

Across all propositions, the analysis suggests that the central explanatory force lies less in modality per se than in the interaction architecture of the examination system: context variability, UI quality, error-consequence regimes, monitoring interventions, and accessibility conditions together form a bundle of outcome-relevant pathways that shape the measurement validity and fairness of digital examinations. The propositions thus function as elements of a causally structured system perspective; their actual function in this article lies in the precise translation of normative quality requirements into empirically operationalizable assumptions of effect, which can guide future research toward a more selective identification of confounding factors in CBT settings.

Discussion

Inductively, the results suggest that the widespread attribution of performance differences to primarily modality-driven mode effects remains analytically underdetermined once digital exams are understood not as a mere change of medium but as a socio-technical interaction system.

Given the research gaps noted here, performance outcomes should be interpreted from this systemic perspective not so much as a direct function of the input technology—namely, keyboard versus handwriting—but rather as the result of interaction-mediated processes that are influenced by the usage context, system quality, the consequences of errors, monitoring mechanisms, and accessibility conditions (Ceka & O'Geen, 2019; Eleanor *et al.,* 2014; Weirich *et al.,* 2024).

From a premise-based perspective, the conclusion implies that inconsistent results in comparisons between CBT and paper should not be interpreted as a contradiction but rather as an expected manifestation of insufficiently controlled variability in the experimental conditions (Al-Fraihat *et al.,* 2020; Al-Qdah & Ababneh, 2017; Ceka & O'Geen, 2019; Dendir & Maxwell, 2020; Eleanor *et al.,* 2014; Erzurumlu, 2023; Ita *et al.,* 2014; Ravesloot *et al.,* 2026; Wang *et al.,* 2008; Weirich *et al.,* 2024): Essentially, this conclusion means that where studies diverge from one another—often due to polysemic contextual factors—this divergence is frequently consistent with the assumption

that different implementation and contextual conditions activate different parts of the mechanisms of action. In this respect, the results of the field studies reviewed in this exploratory study document and discuss the higher scores in the CBT mode. Furthermore, the explanations derived from these findings can be linked, for example, to production and revision advantages associated with expanded responses, without such differences necessarily implying higher construct performance (Deutsches Institut für Normung (DIN), 2020c).

Furthermore, in contrast, various studies showing no score differences but simultaneous differences in preference/acceptance indicate that the similarity of results does not rule out the existence of systematic interaction burdens but merely allows these to become apparent in other variable situations (stress, satisfaction, avoidance behavior) (Browning *et al.,* 2021; Deutsches Institut für Normung (DIN), 2020b; Elmer *et al.,* 2020; Johnson *et al.,* 2020; Raaheim *et al.,* 2019; Vindigni, 2025c; Wagner *et al.,* 2022).

Normative Interpretation: Usability as a Condition of Validity

It can be said that the ISO 9241-based reconstruction ensures a conceptually controlled shift: Usability (DIN EN ISO 9241-11) is therefore not understood as a secondary quality feature of digital exams but as a condition of validity in the sense that effectiveness, efficiency, and satisfaction in the specific context of use help determine the processing actions and thus the construct interpretation of the results (Al-Fraihat *et al.,* 2020; Ceka & O'Geen, 2019; Deutsches Institut für Normung (DIN), 2020b; Geis & Tesch, 2023; Hall *et al.,* 2022; Vindigni, 2026c, 2026b).

Consequently, digital exams can accumulate validity and fairness risks where interaction conditions (e.g., navigation architecture, feedback logic, error handling) are not appropriate to the task or where contextual conditions (terminal/network quality, environmental disturbances) are systematically unevenly distributed. This line of argument is central to the matter at hand because it contextualizes the discussion of performance differences, shifting it from the level of the medium to the level of measurement-relevant interaction conditions. Accordingly, DIN EN ISO 9241-110 functions as a normative translation tool (Deutsches Institut für Normung (DIN), 2020b; Geis & Tesch, 2023; Vindigni, 2025c, 2026b, 2026c): interaction principles such as task appropriateness, self-descriptiveness, expectation conformity, error tolerance, and controllability can be operationalized as exam-related failure modes and thus integrated as verifiable risk indicators in university quality and evaluation processes (Al-Fraihat *et al.,* 2020; Al-Qdah & Ababneh, 2017; Ceka & O'Geen, 2019; Hall *et al.,* 2022; Ita *et al.,* 2014; Weirich *et al.,* 2024).

Finally, DIN EN ISO 9241-210 legitimizes the governance perspective: digital examinations are to be treated as iterative, human-centered systems whose scope of evaluation must be dimensioned in accordance with the risks involved; in this respect, examinations are not just IT but a relevant application domain with potentially serious consequences if requirements are not met (misclassification, exclusion) (Deutsches Institut für Normung (DIN), 2020c; Geis & Tesch, 2023; Lestari, 2024; Vindigni, 2026c, 2026b).

Proctoring and Integrity: Validity as a System Property

This exploratory study also indicates that the academic discussions on digital exams are methodologically incomplete insofar as proctoring and integrity regimes are treated as exogenous constraints. The studies cited in this regard demonstrate that monitoring arrangements can alter performance outcomes and must therefore be considered as an independent path in outcome genesis (Andreou *et al.,* 2021; Cramp *et al.,* 2019; Davis *et al.,* 2016; Meulmeester *et al.,* 2021; Milone *et al.,* 2017; Murphy *et al.,* 2022; Oeding, Gunn, & Jones, 2024).

At the same time, the validity problem of unproctored online summative exams is highlighted by the fact that opportunities for cheating can fundamentally undermine the interpretability of performance data (Newton, 2023). In the author's view, this implies the following for the present system perspective: Validity is not solely a test-theoretical attribute of items or tasks but a property of the entire examination system, including technical, organizational, and social control mechanisms.

However, this conclusion also results in a methodological warning for future comparative studies: CBT vs. paper designs are susceptible to spurious explanations without control or explicit modeling of the monitoring regime (and the stress it induces), because score differences can be mediated by both integrity effects (fraud reduction) and interaction and stress pathways (Ceka & O'Geen, 2019; Oeding, Gunn, & Jones, 2024; Oeding, Gunn, & Seitz, 2024).

Accessibility and Fairness: From Accommodations to Structural Conditions

In this approach, accessibility is not an add-on (accessibility as an extra) but a structural element: Accessibility conditions change resource situations, interaction costs, and error risks—and thus the likelihood that digital exams will assess heterogeneous student groups fairly. The ISO 9241 standards provide a logical connection to this topic via the context of use (9241-11) and the interaction principles (9241-110), while the process standard (9241-210) requires a human-centered, inclusive design and evaluation logic (Al-Qdah & Ababneh, 2017; Andreou *et al.,* 2021; Browning *et al.,* 2021; Ceka & O'Geen, 2019; Deutsches Institut für Normung (DIN), 2018a, 2020b, 2020c; Eleanor *et al.,* 2014; Geis & Tesch, 2023; Ita *et al.,* 2014; Lestari, 2024; Vindigni, 2025c, 2026c; Wagner *et al.,* 2022).

This means that fairness can be interpreted not as a retrospective accommodation decision, but as a design and governance feature of the examination system.

Physical Ergonomics and Anthropometry: Examination Environment as a Measurement-Relevant Condition

As soon as digital exams effectively constitute the physical user environment (workstations, input devices, viewing conditions, body postures), anthropometry becomes a non-trivial boundary condition for equal opportunities. The DIN 33402 series of standards (in particular Part 1: Terms/Measurement Methods; Part 2: Values) provides the reference for determining and using human body measurements and highlights the variability of the user population as a design and evaluation problem (DIN Media GmbH 2020d).

For the context of university examinations, this implies the following: Standardized room and workplace configurations are by no means neutral; they can systematically generate higher interaction costs for subpopulations (e.g., unfavorable eye/hand coordination, unergonomic input devices, unfavorable seat/table geometry) (DIN Media GmbH 2020d). From a system perspective, such effects should not be treated as comfort but as potentially outcome-relevant (DIN Media GmbH 2020d).

In this respect, ISO/TR 9241-514 complements the above perspective as a guideline for the application of anthropometric data within the ISO 9241-500 series and thus addresses the methodological question of how anthropometric variation should be incorporated into the design and evaluation of interactive system environments (DIN Media GmbH 2018a; DIN Media GmbH 2020a).

Furthermore, DIN EN ISO 15537 also frames the selection and use of test subjects for testing anthropometric aspects: For the evaluation of digital testing environments, this leads to a governance implication that panels or pilots should be composed in a way that is appropriate for the population, rather than being opportunistic, if anthropometric and body-related interaction conditions are considered essential for fairness (Deutsches Institut für Normung (DIN), 2022c).

In addition, DIN EN 614-1 can also be used as an overarching ergonomic guideline: even if the machine context is not 1:1 transferable to university examinations, the standard legitimizes the fundamental insight that ergonomics should be treated as a terminologically and conceptually explicit design perspective on interactions for the purposes of praxeology in the tertiary education environment—especially where poor design not only induces efficiency losses but also systematic risk and misclassification consequences (Deutsches Institut für Normung (DIN), 2009).

Research Agenda: More Selective Operationalization Instead of Further Score Comparisons

From this standpoint, the synthesis should therefore result in a research agenda that focuses less on renewed global score comparisons and more on the more selective identification and measurement, i.e., ex ante (conceptual), interim (conceptual), and ex post (results-oriented), of the paths reconstructed in the model (Vindigni, 2025c, 2026b, 2026c). Specifically, this implies that:

a) Context variables (device/network quality, room/environmental conditions, display quality, workplace configuration) must be systematically collected and modeled as moderators (Deutsches Institut für Normung (DIN), 2018a, 2018c, 2018b).

b) Interaction metrics (navigation and editing effort, error events, recoverability, and abortions) must be operationalized as mediating variables (Deutsches Institut für Normung (DIN), 2020b).

c) Proctoring regimes must not only be documented but also taken as a separate effect path (with stress/strain indicators, if applicable).

d) Anthropometric and physical-ergonomic conditions must be examined as fairness conditions where relevant (DIN Media GmbH 2020d; DIN Media GmbH 2020a; DIN Media GmbH 2022c). These criteria must be modeled as system criteria across task, interface, and governance levels (Deutsches Institut für Normung (DIN), 2020c, 2022b; Lestari, 2024; Oeding, Gunn, & Jones, 2024; Oeding, Gunn, & Seitz, 2024).

Limitations of the Exploratory Study and Robustness Argument

Since the present contribution is conceived not as a controlled usability study but as a normatively guided process of model construction and evidence-triangulating reconstruction, its epistemic achievement does not lie in the quantitative estimation of discrete effect magnitudes but rather in the explication and systematic ordering of those causal pathways through which digital examinations, as socio-technical interaction systems, can shape validity, fairness, and performance representation under specific contextual conditions.

However, this approach is linked to arguments:

The robustness claim is grounded in evidence triangulation: integrating comparative studies, proctoring/integrity research, and mechanism-oriented literature reduces the risk of overinterpreting isolated findings; nonetheless, the inference remains contingent on the reporting quality and methodological transparency of the underlying primary studies (Lestari, 2024).

An additional limitation lies in the heterogeneity of the tertiary examination field: subject cultures, task formats, and institutional infrastructures vary considerably, which means that generalizations about digital exams can only be made as statements about constellations (not about individual technologies) (Al-Qdah & Ababneh, 2017; Ceka & O'Geen, 2019; Deutsches Institut für Normung (DIN), 2020c, 2020b; Milone *et al.,* 2017; Newton, 2023; Noorbehbahani *et al.,* 2022; Oeding, Gunn, & Jones, 2024).

Core Implication: Design, Evaluation, and Governance as a Necessary Consequence

Taken together, the findings indicate that universities

must treat digital examinations as socio-technical examination systems whose governance cannot be reduced to platform choice or mere integrity controls. Accordingly, the analysis yields (i) normatively grounded test criteria, (ii) ISO 9241-110–compliant design requirements that translate dialogue quality into exam-relevant validity and fairness risks, and (iii) minimum measures to systematically reduce interaction load, error-consequence exposure, and exclusion—embedded in a human-centered evaluation and continuous improvement cycle in line with ISO 9241-210.(Deutsches Institut für Normung (DIN), 2020c, 2020b).

Also, extended standard references specify what these minimum measures refer to beyond UI design: environmental and equipment ergonomics (ISO 9241-500), display and viewing conditions (ISO 9241-306), and—if test settings allow physical interaction costs to vary systematically—anthropometric principles and their evaluative handling (DIN Media GmbH 2018a; DIN Media GmbH 2018c; DIN Media GmbH 2020d; DIN Media GmbH 2020a; DIN Media GmbH 2022c). In certain usage contexts, the justified provision of analog alternatives is also objectively compatible as a governance-capable option, not as a step backward, but as part of a validity- and fairness-oriented testing architecture, provided that digital constellations do not guarantee the risk-adequate fulfillment of the aforementioned conditions.

Practical Implications and Governance: Normative Test Criteria, Design Requirements, and Minimum Measures

In substantive terms, the present findings entail a governance-relevant consequence for tertiary education: digital exams should not be treated as a mere administrative variant but as relevant socio-technical systems whose outcome profiles are conditioned by interaction conditions, error consequences, monitoring regimes, and context variability, because these factors can materially shape the conditions of valid and fair assessment.

This shifts the question of responsibility from individual competence in dealing with technology to the institutional obligation to provide an exam arrangement that is effective, efficient, and satisfactory for use in defined contexts of use in accordance with DIN EN ISO 9241-11 (Deutsches Institut für Normung (DIN), 2018a). At the same time, interaction principles according to DIN EN ISO 9241-110 must be operationalized as test-related quality and risk criteria and transferred to a human-centered design and evaluation process in accordance with DIN EN ISO 9241-210 (Deutsches Institut für Normung (DIN), 2020c, 2020b).

Normative Minimum Requirements as an Examination Architecture (MVP logic)

Against the backdrop of limited resources, a minimum viable policy (MVP) should not be understood as minimalism; it instead represents a minimum architecture, defined as a set of non-negotiable minimum measures that must be fulfilled in all digital examinations to systematically reduce validity and fairness risks.

Minimum architecture (university-wide):

a) Defined context of use (target population, device policy, room/remote conditions, support processes) as part of the test documentation (DIN EN ISO 9241-11) (Deutsches Institut für Normung (DIN), 2018a).

b) Principle-based system requirements (DIN EN ISO 9241-110) as exam-related quality criteria (e.g., error tolerance, conformity with expectations, task appropriateness, controllability) (Deutsches Institut für Normung (DIN), 2020b).

c) Risk-adequate evaluation and follow-up (DIN EN ISO 9241-210): formative pilots, summative acceptance, continuous field observation, and incident-driven improvement (Deutsches Institut für Normung (DIN), 2020c).

d) Accessibility as an integral part of the test architecture (DIN EN ISO 9241-20)—not as a subsequent correction in individual cases (Deutsches Institut für Normung (DIN), 2022b).

e) Environment and display quality as a contextual condition (ISO 9241-500; ISO 9241-306)—especially for BYOD and variable room settings (Deutsches Institut für Normung (DIN), 2018c, 2018b).

f) Anthropometric basis: Universities are required to provide exam rooms and workstations that are anthropometrically specified and adjustable in accordance with DIN 33402, ISO/TR 9241-514, and DIN EN ISO 15537, thereby preventing systematic exclusion effects and ensuring that physical interaction constraints cannot distort exam performance as a function of body-related fit rather than competence. (DIN Media GmbH 2020d; DIN Media GmbH 2020a; DIN Media GmbH 2022c).

ISO/DIN-Based Test Criteria Matrix (compact, operational, auditable)

The following matrix is intended as an auditable

Table 3: Normatively derived criteria and minimum measures to mitigate validity, fairness, and outcome risks in digital exams

standard_ anchor	exam_ relevant_focus	typical_failure_ mode	risk (validity/ fairness/outcome)	minimum_ measure (mvp)	minimal_ indicator
din en iso 9241-11	specify context of use	undefined context (devices, network, room/remote conditions, support)	outcome variance driven by context variability; unfairness	context-of-use sheet per exam; device policy + support process	documentation completeness; incident rate

din en iso 9241-110	suitability for the task	navigation forces split attention	efficiency loss; errors; increased time-on-task	single-screen logic where possible; clear structure; reduced click paths	click-path length; scroll depth; completion time
din en iso 9241-110	error tolerance	irreversible text loss / incorrect submission	system-induced performance loss; unfairness	autosave; versioning; confirm-submit; recovery options	rate of irreversible errors; recovery success rate
din en iso 9241-110	self-descriptiveness	unclear status/error communication	stress/strain; error accumulation	unambiguous status indicators; actionable error messages; clear next steps	help requests; error loops
din en iso 9241-110	conformity with user expectations	inconsistent editor/shortcuts	disadvantages for specific routines; avoidable friction	standardized editors; training mode; consistent shortcuts	training completion; usability complaints
din en iso 9241-110	controllability	lockdown/ proctoring reduces user control	stress; aborts; performance drop	clear control elements; defined escape/assist pathways	abort rate; proctoring interventions
din en iso 9241-210	human-centred design process	go-live without pilots/acceptance evaluation	latent defects → outcome artefacts	pilot + summative acceptance; incident review cycle	pilot report; post-exam incident log
din en iso 9241-20	integrated accessibility	assistive technology breaks the workflow	exclusion; systematic disadvantage	accessible patterns; testing with assistive technology	accessibility test report; accommodation success
iso 9241-500	environment/ workstation ergonomics	inadequate desk/ seat/input setup	strain; errors; fairness risk	standard workstations; ergonomics check	ergonomics checklist; complaint rate
iso 9241-306	display quality	glare/contrast issues	reading errors; time loss	display & lighting check; minimum requirements	readability check; display-related incidents
din 33402 / iso/tr 9241-514	anthropometry	one-size-fits-all workstations	systematic interaction costs	workstation variants; anthropometric verification	fit rate; posture issues
din en iso 15537	test participant selection	non-representative pilots	risks remain undetected	vary panels according to the target population	panel composition compliance
din en 614-1 / din en iso 26800	ergonomic principles	no systematic ergonomics/ governance integration	governance blind spots	ergonomics principles embedded in policy	audit finding rate

Source: Author (2026)

transformation grid: it translates standard principles into practical failure modes, risks, and minimum measures for testing. (The operationalizations are chosen so that they can actually be collected in university IT governance and quality assurance.)

Annotated Reasoning Logic

This matrix operationalizes the paths reconstructed in this exploratory study (interaction load, error consequences, monitoring, and context variance) as verifiable minimum requirements. It can therefore be linked to university QM, IT security governance, and examination regulations without reducing the standards to mere checklist rhetoric.

Proctoring Governance: Ensuring Integrity without Compromising Validity/Fairness

Across the proctoring and assessment integrity literature, monitoring regimes are reconstructed as non-neutral components of the testing system that can shape performance via strain, behavioral adaptation, and disruption pathways. Accordingly, governance frameworks must not treat proctoring as a purely integrity-preserving add-on but as an intervention whose design parameters and error/flag regimes carry validity and fairness implications. Moreover, from a governance theory perspective, this is consistent, as digital infrastructures regularly form new intermediation and gatekeeping structures through which responsibility, control points, and competence burdens are shifted from the system to individuals—a pattern that can also be reconstructed in tertiary education infrastructures (e.g., credential/identity ecosystems) as a re-intermediation mechanism (Vindigni, 2025e, 2025a, 2025f, 2025d, 2025b, 2025c, 2026b, 2026a.

An appropriate governance approach therefore involves a double check:

a) Integrity check: Which risk situation should be addressed (exam stakes, opportunities for cheating, item exposure)?

b) Interaction/stress check: What additional interaction costs, consequences of errors, privacy, and stress effects does proctoring induce (including setup complexity, false positives, and loss of control)?

Minimum measures (proctoring-specific):

• Proctoring only according to risk classification (low/medium/high stakes) and as a documented decision (ISO 9241-210) (Deutsches Institut für Normung (DIN), 2020c).

• An assist pathway: defined escalation and assistance processes without punitive automatisms.

• Transparent status/rule communication (ISO 9241-110: self-descriptiveness, conformity to expectations) (Deutsches Institut für Normung (DIN), 2020b).

• Logging as a learning tool: post-exam incident review (false positives, technical interruptions) for iterative improvement (ISO 9241-210) (Deutsches Institut für Normung (DIN), 2020c).

Analog Alternatives as a Risk-Appropriate Option (not as a counterprogram)

From a system perspective, the option of analog alternatives should not be justified as nostalgia; it should instead be framed as risk mitigation in critical usage contexts: Where digital constellations (device heterogeneity, unstable networks, uncontrollable error risks, lack of accessibility certification) do not sufficiently meet the requirements for validity and fairness, the justified provision of analog testing methods can be a governance-rational measure.

However, this option is not absolute but contextualized: it must be linked to documented criteria (e.g., uncontrollable error risks, lack of accessible usage contexts, and unfulfillable display/environmental requirements).

Accessibility as a governance and compliance regime since June 28, 2025 (BFSG/EAA, EN 301 549, WCAG/BITV)

On June 28, 2025, a regulatory threshold was reached in the European context: the accessibility requirements under the European Accessibility Act (Directive (EU) 2019/882) became applicable throughout the European Union, while in Germany the Barrier-Free Accessibility Act (BFSG) became binding as the national implementation framework for certain products and services, including digital services (Bundesministerium für Digitales und Staatsmodernisierung, 2025b; European Union, 2019). For universities—especially those where exams are provided and conducted digitally (CBT settings, remote exams, proctored formats)—accessibility is thus shifting from a primarily ethical and didactic quality requirement to a minimum condition relevant to exam governance, which directly contributes to assumptions of validity and fairness.

Analytically, we must emphasize that WCAG compliance alone cannot adequately define accessibility in practice (W3C Web Accessibility Initiative, 2025). Rather, EN 301 549 forms the authoritative operational reference framework for accessibility requirements for ICT products and services in Europe and acts as a technical translation standard between legal obligations and verifiable system design (Deutsches Institut für Normung (DIN), 2022a). In the German administrative and higher education context, this logic must also be systematically intertwined with BITV 2.0 as the implementation framework in the public sector (Bundesministerium für Digitales und Staatsmodernisierung, 2025a). For digital exams, this means that accessibility should not be treated as an add-on to the user interface but as an end-to-end feature of the entire exam service (access paths, authentication, editor/response formats, navigation, feedback/status, error tolerance, proctoring layer, and support paths).

In light of this, the ISO 9241-based system perspective, as developed here, yields two governance implications: First, digital exams must be audited as compliance-critical interaction systems, with EN 301 549/BITV/WCAG defining the formal requirements as a test grid, while the ISO 9241 logic explains the outcome-relevant usability and interaction risks (Bundesministerium für Digitales und Staatsmodernisierung, 2025a; Deutsches Institut für Normung (DIN), 2022a; European Union, 2019; W3C Web Accessibility Initiative, 2025). Second, the widespread implementation practice cannot be empirically assumed to be sufficiently reliable. In a relevant systematic research review presented by the author, it is shown that, despite normative consolidation (WCAG/BITV, EAA/BFSG, EN 301 549), institutional implementation often fails due to fragmented responsibilities, a lack of QA routines, and insufficient enforcement. Accessibility is thus formally asserted but not stably realized as a participation infrastructure (Bundesministerium für Digitales und

Staatsmodernisierung, 2025a; Deutsches Institut für Normung (DIN), 2022a; European Union, 2019; Vindigni, 2026b; W3C Web Accessibility Initiative, 2025). In this regard, for the governance of digital examinations, this has a pragmatically harsh implication: Where accessibility certificates (EN 301 549/BITV/WCAG) are lacking or are not continuously ensured during operation, there is not only a risk of exclusion but also potentially examination-related and reputational vulnerabilities (fairness/equal opportunities)—especially in remote settings where students may be dependent on assistive technologies, alternative input paths, or support due to their context (Bundesministerium für Digitales und Staatsmodernisierung, 2025a; Deutsches Institut für Normung (DIN), 2022a; European Union, 2019; W3C Web Accessibility Initiative, 2025).

Implementation Path for Universities

However, to ensure that the minimum requirements outlined above do not remain merely normative but can be translated into an institutionally robust examination practice, a pragmatic yet normatively binding implementation path is recommended. Such a path should not be understood as a linear technical or even mundane introduction, but rather as a governance-led process in which digital examinations are institutionally stabilized as socio-technical examination systems within defined contexts of use.

Policy Set and Responsibility Architecture

First and foremost, the university needs a binding regulatory framework that explicitly defines the specific context in which digital exams are used and translates the developed minimum architecture into university-wide examination guidelines, structures of responsibility, and procedural safeguards. Such a framework is not merely advisable for reasons of organizational consistency but is particularly necessary from a normative perspective because digital exams in the tertiary sector—especially where distance learning courses are offered digitally and exam performance is conveyed, delivered, or monitored accordingly—are subject to heightened requirements regarding accessibility, equitable participation, and institutional responsibility. Against this backdrop, requirements from the BFSG, the European Accessibility Act, the BITV, the WCAG, and EN 301 549 cannot be treated as peripheral compliance matters; rather, they necessitate that usage contexts, technical and situational access conditions, potential barriers, support and escalation channels, as well as justified alternative arrangements, be determined in advance and in a regulated manner (Bundesministerium für Digitales und Staatsmodernisierung, 2025a; Deutsches Institut für Normung (DIN), 2022a; European Union, 2019; W3C Web Accessibility Initiative, 2025). The usage context is thus not merely an empirical backdrop but, itself, the subject of normative definition and institutional accountability. This includes, at a minimum, the definition of relevant target groups; assignment and exam formats; device and access requirements; on-campus and at-home usage contexts; support and escalation channels; and accessibility requirements, as well as the clear allocation of procedural responsibilities among the exam administration, IT, accessibility teams, and faculty. The usage context is thus no longer treated as an implicit background assumption but becomes an integral part of the auditable examination architecture.

Risk-Appropriate Pilot Testing Under Realistic Conditions

Before operational implementation into regular examination procedures, a risk-appropriate testing phase must be conducted, during which each examination platform, along with the relevant task and interaction formats, is tested under conditions that realistically reflect the intended usage context in terms of its technical, situational, and workload-related complexity. In accordance with DIN EN ISO 15537, the pilot panel should be sufficiently representative, i.e., it should take into account different usage scenarios, prior experience, technical requirements, and accessibility needs (Deutsches Institut für Normung (DIN), 2022c). The pilot testing must not be limited to a mere functional test but must systematically document interaction effort, error events, recoverability, support needs, contextual disruptions, and potential risks of exclusion. This particular reconstruction demonstrates that outcome-relevant problems in digital exams often do not arise solely from system failures in the narrow sense, but rather from the interplay of interaction conditions, contextual variability, and the consequences of errors.

Summative Approval and Proof of Accessibility

Before being transferred to regular examination operations, a normatively grounded and institutionally accountable evaluation decision is required, which must not be limited to the formal approval of a technically functional system. Rather, it must be examined whether the specific digital examination setup, within the intended context of use, sufficiently possesses the characteristics that qualify it as a valid, fair, controllable, fault-tolerant, and accessible examination environment. This assessment must be guided by the interaction principles of DIN EN ISO 9241-110, the accessibility-related requirements of ISO 9241-20, and the relevant normative and legal obligations for barrier-free digital examination environments (Deutsches Institut für Normung (DIN), 2020b, 2022b). The subject of the evaluation is therefore not merely technical operability, but the institutionally accountable question of whether the specific examination architecture ensures, under the conditions of its actual use, a level of validity, fairness, accessibility, and equitable participation that is commensurate with the risks involved. Where such criteria cannot be demonstrated with sufficient certainty and justification, the implementation lacks a normative basis for justification; it must then be revised and re-evaluated prior to implementation.

Operational Implementation with Established Support and Fallback Procedures

The operational rollout should also be carried out within the university's organizational framework, in conjunction with clearly defined support, incident, and fallback procedures. These include documented contact and escalation procedures during the audit and transparent procedures for technical malfunctions and recovery options to the extent they are objectively feasible, as well as justified analog or alternative arrangements in critical usage scenarios where digital conditions do not ensure the risk-appropriate safeguarding of validity, fairness, and accessibility. In this respect, analog fallback architectures should not be understood as a failure of digitalization but rather as an integral part of a responsible audit architecture.

Continuous Field Monitoring and Incident-Based Adjustments

Finally, digital examination systems must be subject to continuous field monitoring in accordance with the process logic of DIN EN ISO 9241-210. After each testing phase, incidents, accessibility-related impairments, support cases, error patterns, and indications of systematic strain or exclusion should be evaluated in a structured review cycle and fed back into subsequent design, evaluation, and governance decisions. Digital exams thus do not appear as finished products, rather as socio-technical arrangements that require ongoing observation, evaluation, and risk-appropriate adaptation (Deutsches Institut für Normung (DIN), 2020c).

Figure 3 below illustrates the integrated normative pathway for digital exams developed in this study. In this context, DIN EN ISO 9241-11 defines usability as context-dependent, in this case learner-centered, goal achievement in terms of effectiveness, efficiency, and satisfaction, while DIN EN ISO 9241-110 translates this foundation into an operational logic of heuristic complexity reduction through concrete interaction principles Deutsches Institut für Normung (DIN), 2018a; Deutsches Institut für Normung (DIN), 2020b; Geis & Tesch, 2023; Vindigni, 2026c). Last but not least, DIN EN ISO 9241-210, which normatively specifies the human-centered design of interactive systems, frames the human-centered governance and improvement process within which learner-centered digital examination systems can be implemented and further developed iteratively, responsibly, and as institutionally robust arrangements

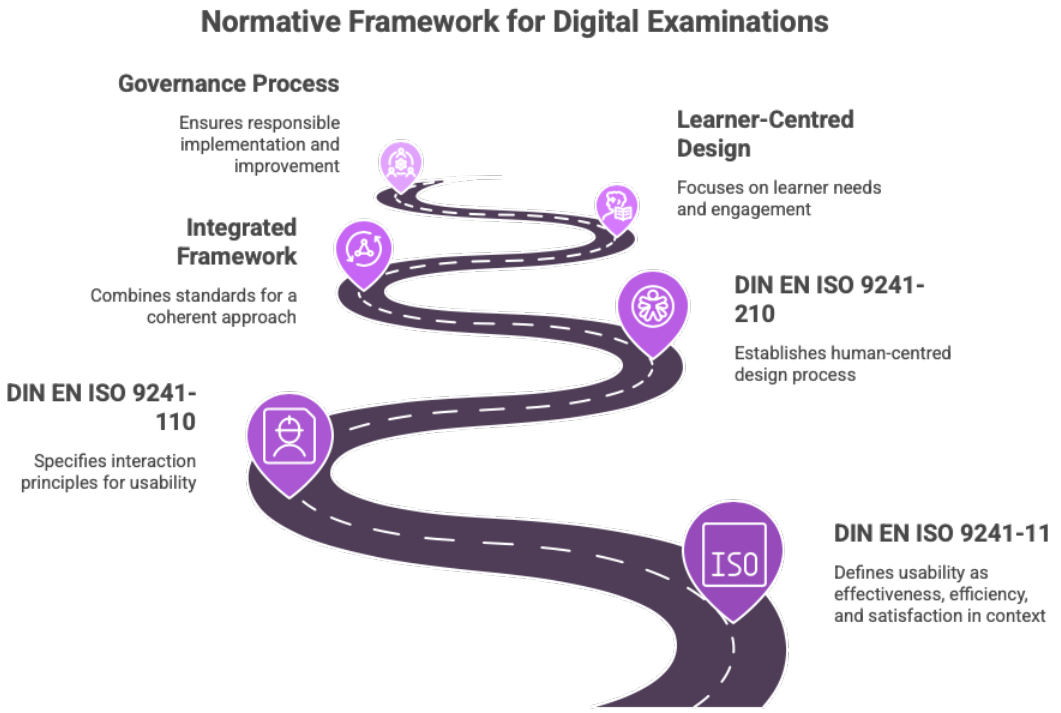

Figure 3: Integrated relevance of DIN EN ISO 9241-11, -110, and -210 for learner-centred digital examinations
Source: Author (2026)

under real-world conditions of use (Deutsches Institut für Normung (DIN), 2020c; Geis & Tesch, 2023; Vindigni, 2026c).

From this perspective, higher education institutions cannot reduce implementation to the mere introduction of a platform. Rather, implementation requires the institutionalization of a context-sensitive, accessibility-oriented, and risk-appropriate governance process that systematically links normative standards, evaluation procedures, and operational responsibilities. This is all the more important because the interaction principles of DIN EN ISO 9241-110 cited in this study are not to be understood merely as criteria of technical or formal usability (Deutsches Institut für Normung (Deutsches Institut für Normung (DIN), 2020b, 2022b).

Within the scope reconstructed here, they instead

contribute to a learner-centered reduction of unnecessary situational and interactional complexity in the digital environment and thereby create conditions under which orientation, controllability, error manageability, adaptability to individual needs, and learning support can become effective in the first place in ways that enable focused, self-regulated, and less stressful exam-taking. This argument is particularly significant because, from the perspective of instructional psychology, learning and taxonomic processes of acquisition cannot be reduced to isolated cognitive operations but are always co-determined by perception, stress, attribution of control, interactional security, and situational adaptability. Only where task appropriateness, self-descriptiveness, error tolerance, controllability, adaptability, and learning support are taken seriously in their interplay can digital exam design be shaped in such a way that participation, inclusivity, and ultimately the democratization claim advanced in this study are not merely asserted programmatically but can be realized institutionally and evaluatively.

It should further be noted that, in the day-to-day institutional routine of digital teaching, learning, and assessment scenarios, it is not uncommon to lose sight of whom these arrangements are ultimately meant to serve: namely students, who are expected to become, or remain, capable of participating in higher education in ways that are adaptive, innovative, and future-oriented. Against this background, fairness toward the target group cannot be reduced to downstream outcome monitoring alone. Whether digital examination and learning architectures fulfill their purpose is not determined ex post by quantified results alone, but already ex ante by their conceptual alignment with objectives and, in the interim, by the quality of their procedural realization. It is precisely here that the particular significance of DIN EN ISO 9241-11 lies (Deutsches Institut für Normung (DIN), 2018a): wherever digital scenarios are designed for instructional or examination purposes, usability must be understood not merely as operational capability but as a context-dependent unity of effectiveness, efficiency, and satisfaction. This satisfaction should by no means be misunderstood as a mere measure of convenience; rather, from a learner-centered perspective, it addresses the question of whether an arrangement is in fact manageable and meaningfully accessible for the target group in a way that enables orientation, participation, appropriate workload, and sustainable academic connectedness. Only where this perspective is taken seriously can one speak of genuine fairness toward the target group in the sense of the democratization discourse addressed here. Higher education governance must therefore not be narrowed to ex-post performance indicators or to mere logics of economic efficiency or reach but must treat conceptual review, process-accompanying evaluation, and outcome assessment as interrelated levels of a learner-centered and institutionally accountable quality architecture.

Against this background, the present article further suggests, from a learning-taxonomic perspective, that independent scholarly and institutional contributions in the field of digital examinations can become viable and sustainably consequential only where certain core concepts are not merely named but understood and internalized in terms of their normative and analytical scope. These include the reconstruction of digital examinations as socio-technical systems, the determination of usability as a condition of validity, the translation of interaction principles according to DIN EN ISO 9241 into concrete categories of fairness and risk, the classification of proctoring as an independent regime of influence and monitoring, and the consideration of accessibility and contextual conditions as examination-relevant variables. What matters here is not merely declarative knowledge of these concepts, but the ability to apply these categories to concrete examination constellations, to reconstruct their pathways of influence, and to assess independently their consequences for the validity, fairness, and governance of digital examinations. In this respect, the article not only offers a systematic reconstruction but also provides a conceptual and normative foundation upon which further research and institutional practice can build. This is all the more relevant as higher education institutions are increasingly required to align their examination and quality-development processes with broader standards of responsible, inclusive, and sustainable development. The categories elaborated here can therefore also be read in light of the Sustainable Development Goals and the Standards and Guidelines for Quality Assurance in the European Higher Education Area, insofar as they point to equitable participation, accessibility, responsibility, quality development, and institutional accountability.

Applied Implications for Socio-Technical Exam Systems in Higher Education

Building on the ISO-9241-based reconstruction of socio-technical impact pathways, the following section illustrates how the proposed framework can be translated into analytically grounded application scenarios for tertiary examination contexts, thereby rendering governance, accessibility, and interaction-sensitive design requirements operational.

To illustrate the practical applicability of the proposed framework model, it seems appropriate to apply the previously developed concepts to specific contexts of higher education examination practice. The aim of the following case vignettes is not to present singular individual cases but to heuristically explain typical scenarios in which digital exams, as socio-technical interaction systems, may present specific risks regarding validity, fairness, and exclusion. It is assumed that digital examination formats should not be analyzed in isolation as technical implementations but primarily as integrative arrangements comprising interaction design, contextual conditions, regulatory requirements, and learning-psychological mechanisms. In this respect, the case

vignettes serve to concretize the previously developed framework in a practical sense and illustrate how the normative reference categories of DIN EN ISO 9241, in conjunction with digital accessibility requirements, can be translated into concrete design decisions.

Case Vignette: Synchronous Digital Exam Environments (Zoom-based empirically instructive reference case)

A first—and, from a methodological perspective, particularly instructive—application arises from synchronous digital teaching, learning, and assessment scenarios, such as those found in video-based online settings with a high level of interaction (Vindigni, 2025e). An empirical study conducted by the author in a higher education context, which combines objective usage data from the Zoom platform, psychometric stress indicators (DASS-21), and a standardized measurement of inhibitory attentional control using an antisaccade task, shows that longer periods of digital exposure tend to be associated with increased stress levels, while compensatory moderation through inhibitory control mechanisms could not be demonstrated (Vindigni, 2025e). At the same time, it becomes clear that digital stress cannot be understood as a mere consequence of screen time; it should instead be seen as the result of a multifaceted interplay between duration of use, interaction architecture, task structure, and situational contextual conditions (Vindigni, 2025e). Thus, this study offers a particularly suitable case study to illustrate the idea developed here that digital examination and performance situations should not be analyzed as neutral media transitions nor as independent socio-technical stress and interaction arrangements (Vindigni, 2025e).

Of particular relevance in this context is that the aforementioned study does not stop at a purely descriptive observation of correlations but explicitly classifies the findings within the normative framework of DIN EN ISO 9241-11, 9241-110, and 9241-210 (Deutsches Institut für Normung (DIN), 2018a, 2020b, 2020c). The logic reconstructed therein shows that psychological stress in digital settings depends significantly on whether interaction and dialogue structures are designed to be task-appropriate, controllable, in line with expectations, and cognitively unburdening. In this regard and in particular, breakout phases, activation-intensive communication formats, and condensed phases of synchronous cooperation appear as constellations of heightened cognitive and social strain, in which deficits in interaction design can have a particularly significant impact (Vindigni, 2025e). This insight is relevant to the examination perspective of interest here because digital exams often exhibit the same stress-inducing characteristics identified in the study: time pressure, high demands on attention, reduced tolerance for errors, technical dependencies, and the need to manage operational tasks and performance simultaneously (Vindigni, 2025e).

In light of these research explorative findings, the case vignette can be summarized as follows: If mild but significant stress effects can be observed in digital teaching situations under realistic conditions, depending on usage intensity and design quality, then it follows that in digital examination situations, factors such as interaction quality, task format, navigation logic, feedback structure, technical robustness, and consequences of errors are not merely peripheral usability aspects, but rather decisive factors in determining whether performance can be achieved in a valid, fair manner and without additional task-relevant stressors (Vindigni, 2025e). The author's empirical study is thus not merely a parallel finding from a related field but serves as practical evidence that the learner-centered and human-centered design of digital systems required by DIN EN ISO 9241 becomes particularly relevant to examinations where high cognitive load and low tolerance for errors converge. In this respect, the case confirms the basic assumption of this paper, namely that digital exams must be treated as interaction systems whose outcome quality is co-constituted by design conditions (Vindigni, 2025e).

For universities, the finding implies in practical terms that synchronous digital exam architectures must not be introduced solely based on considerations of technical provision or administrative scalability (Vindigni, 2025e). Rather, a systematic preliminary assessment of the respective examination setup is required with regard to cognitive loads, interaction logic, error tolerance, peak loads, and support structures (Vindigni, 2025e). The added value of this case vignette thus lies in the fact that it makes visible the transition from general normative modeling to an empirically applicable implementation (Vindigni, 2025e). Where these prerequisites cannot be sufficiently guaranteed, the provision of alternative or compensatory arrangements is also appropriate from a governance perspective because responsibility must not then rest with the students' individual ability to adapt but must be borne institutionally (Vindigni, 2025e).

The added value of this case vignette thus lies in the fact that it makes visible the transition from general normative modeling to an empirically applicable application. It demonstrates that issues of usability, dialogue design, cognitive load reduction, and human-centered system development are not merely downstream optimization options for digital exams but must already be understood at the level of exam architecture as prerequisites for valid, fair, and inclusive performance assessment (Vindigni, 2025e). Precisely because the study in question also draws on real-world usage contexts, objective log data, and a multimodal measurement architecture, it is particularly well-suited to substantiating the argument developed here not only theoretically but also in terms of practical application.

Accessibility and Adaptability Considerations in Digital Exam Environments

Not only can the empirical study presented above be interpreted in terms of cognitive load dynamics, but it also allows for a nuanced examination of those contextual and interactional conditions that, in accordance with DIN EN ISO 9241, can be understood as prerequisites

for adaptive, learner-centered, and potentially barrier-free use (Vindigni, 2025e). Precisely the importance of interaction density, communication structure, the sequencing of collaborative phases, and varying attention demands—as revealed in the study—makes it clear that digital systems do not passively accommodate different usage requirements but can, depending on their design, either amplify or compensate for them (Vindigni, 2025e). From this perspective, accessibility appears not merely as a technical-formal property of a user interface but as the result of the alignment between usage context, task structure, interaction logic, and individual action requirements (Vindigni, 2025e).

In fact, the author's study is especially instructive in this regard because it explicitly links the issue of psychological stress to the quality of digital learning environments and their normative classification in DIN EN ISO 9241-11, DIN EN ISO 9241-110, and DIN EN ISO 9241-210 (Deutsches Institut für Normung (DIN), 2018a, 2020b, 2020c). Even the design of the study makes it clear that digital interaction conditions should not be viewed in isolation as operational phenomena; they should instead be seen as contextual factors relevant to learning and stress. The study does not merely examine whether digital exposure generates stress but specifically explores under what conditions—duration of use, interaction quality, and learner-centered design—interact and how these relationships should be interpreted in light of usability, human-centered design, and accessibility requirements (Vindigni, 2025e). Thus, even in its fundamental epistemological structure, the study addresses the adaptability, stress sensitivity, and design flexibility of digital systems, and precisely for this reason, it builds a solid bridge to the issue of accessibility.

However, from the perspective relevant here, it is particularly significant that the author's empirical findings do not attribute the cognitive load of digital teaching-and-learning environments solely to the duration of use, as this would be a monocausal explanation, instead attributing it to qualitative design features of the environment (Vindigni, 2025e). In this context, breakout sessions are highlighted above all as phases of heightened cognitive and social activation, during which the demands on orientation, self-regulation, communicative coordination, and situational adaptability noticeably increase (Vindigni, 2025e). These very constellations are also relevant for digital examination environments: where intensified interaction, time pressure, reduced error tolerance, and increased demands on attention converge, the question becomes more acute as to whether the examination environment actually accommodates diverse usage requirements or whether it generates exclusionary and overwhelming effects due to contextual factors (Vindigni, 2025e). In this respect, the study makes it clear that it is not only manifest technical defects but also an inadequately calibrated interaction architecture that can give rise to participation risks.

Similarly, with regard to DIN EN ISO 9241-11, it must be noted that, from a praxeological perspective and taking into account implications from instructional psychology and instructional design, usability is not reduced to mere user-friendliness but is understood as the interplay of effectiveness, efficiency, and satisfaction within the specific context of use (Deutsches Institut für Normung (DIN), 2018a). When transferred and applied to digital exams in the tertiary education context, this normative definition means that an exam environment is not sufficiently suitable merely because it is technically accessible and functionally operational; rather, it must be designed in such a way that exam performance can be achieved under real-world conditions without unnecessary additional stress, without avoidable disorientation, and without systemic disadvantages (Vindigni, 2025e). In this respect, the findings highlighted in the study regarding the importance of clear instructions, structured activation, tools suitable for everyday use, and interaction designs that reduce psychological stress can therefore readily be interpreted as requirements for digital examination systems. In other words, where such requirements are not met, accessibility risks becoming, in practice, a mere formal designation without being realized as actual openness of use and the ability to participate (Vindigni, 2025e).

Moreover, DIN EN ISO 9241-110 provides a normative framework based on the principles of task appropriateness, self-descriptiveness, controllability, conformity to expectations, error tolerance, and customizability, which can be used to more precisely classify the stress and adaptation problems observed in the study (Deutsches Institut für Normung (DIN), 2020b). For example, if breakout phases or other interactive sequences can only be managed under increased stress, even though no isolated protective effect of inhibitory control is detectable, this suggests that stress regulation must be sought not primarily in the individual, but in the design of the system (Vindigni, 2025e). Herein lies the barrier- and adaptation-sensitive insight of the case vignette: Digital examination environments must not be designed in such a way that they tacitly assume a high degree of self-compensation, implicit operational confidence, or uninterrupted concentration stability. Rather, they must be structured so that different rhythms of action, modes of perception, attention profiles, and support needs are anticipated and accommodated within the system itself (Vindigni, 2025e).

In this regard, the author's study further suggests that accessibility should be explicitly understood as an integrative interface between information architecture, interaction logic, and psychosocial support (Vindigni, 2025e). This is already evident from the fact that, in addition to technical usability, clear instructions, gradual activation, appropriate tools, and consistent learning guidance are specifically identified as stress-reducing factors (Vindigni, 2025e). However, such features are not only didactically advantageous but also constitute core criteria of adaptive accessibility: they reduce room for interpretation, minimize operational uncertainties,

facilitate orientation, and lower the likelihood that system-related additional requirements will overshadow the actual performance. For digital exams, the implication is that accessibility does not begin only where assistive technologies are needed; it instead starts where the exam architecture organizes task completion, navigation, feedback, error correction, and time orientation in such a way that different user requirements do not translate into hidden exam disadvantages (Vindigni, 2025e).

In terms of normative synthesis, it can therefore be concluded that the learner-centered and human-centered design logic emerging from the study aligns directly with current requirements for digital accessibility. The author himself explicitly links the increased relevance of adaptive UX design in breakout-based and synchronous online settings to inclusion and digital accessibility as defined by the BFSG, BITV 2.0, and WCAG 2.2 (Bundesministerium für Digitales und Staatsmodernisierung, 2025a; Deutsches Institut für Normung (DIN), 2022a; European Union, 2019; Vindigni, 2026b; W3C Web Accessibility Initiative, 2025). Thus, it also becomes clear that accessibility compliance must not be assessed ex post via legal checklists but must already be embedded in the didactic and ergonomic design of digital interaction environments (Vindigni, 2025e). For digital exams, this requirement implies a clear shift in perspective: It is not the individual who must adapt to a rigid exam environment; rather, the environment must be designed in such a way that it structurally accounts for heterogeneous resource situations, variable usage contexts, and differing support needs (Vindigni, 2025e).

In academic practice, this means that digital exams, which are interaction systems sensitive to accessibility and adaptability, must undergo suitability testing before regular use (Vindigni, 2025e). The testing must cover not only minimum technical functions but also comprehensibility, navigation logic, error handling, the ability to pause and resume, cognitive pacing, alternative paths, and compatibility with different forms of information processing. The study's finding that stress and strain are significantly influenced by the quality of the learning environment necessitates an institutional response: accessibility in the examination context should not be treated as an add-on service or a form of retroactive compensation, but rather as an upstream design and governance task (Vindigni, 2025e). Where such work is not done, there is a risk that digital exams, while formally available, are in fact only manageable under the condition of increased self-compensation or tacit adaptation efforts, which can lead to inequities among students who may not have the same resources or support to navigate these challenges (Vindigni, 2025e).

Furthermore, his case study is valuable because it interprets the empirical study as more than just findings on stress and cognitive control or evidence that design quality, openness to use, and sensitivity to participation in digital environments are inextricably intertwined (Vindigni, 2025e).

It can be concluded that, in the context of higher education pedagogy and digital exams, accessibility should not be viewed as a marginal special requirement in and of itself but primarily as both a structural and normative prerequisite for valid and fair performance assessment (Vindigni, 2025e). In this regard, the author's case description above confirms the central thesis of this article: Digital exams can only be assessed appropriately if their interaction architecture, their context-dependence, and their potential risks of exclusion are systematically and politically incorporated into the evaluation: ex ante (in the sense of a design review), interim (in the sense of a process review), and ex post (in the sense of an outcome review) (Vindigni, 2025e).

Governance-Sensitive Examination Architectures and Institutional Responsibility for Digital Examination Systems

Within the above framework, the case vignette presented earlier demonstrates that digital exams cannot be understood merely as a technical form of distribution for traditional exam formats or as isolated pedagogical decisions; they instead constitute complex socio-technical interaction systems whose design quality is directly linked to questions of validity, fairness, participation, and stress sensitivity (Vindigni, 2025e). Taking this perspective seriously necessitates an expansion of the analysis to encompass governance. Responsibility for the appropriateness of digital examination conditions cannot be shifted to students' individual adaptability but must be understood as an institutional design task (Vindigni, 2025e). In this respect, digital examination environments are not merely pedagogical tools; they instead constitute components of organizational decision-making architectures in which normative requirements, technical infrastructures, and didactic objectives must be harmonized (Vindigni, 2025e).

To this extent, the author's study demonstrates, in an exemplary and equally praxeologically complementary manner, that stress and interaction dynamics cannot be explained solely by individual dispositions but are shaped to a significant degree by the quality of the respective learning and usage contexts (Vindigni, 2025e). However, the convergence of synchronous interaction formats, increased attention demands, and reduced error tolerance clearly shows that the organizational scope for design can influence whether digital arrangements have a compensatory, neutral, or stress-intensifying effect (Vindigni, 2025e). Applying this insight to examination contexts reveals a structural analogy: Here, too, interaction architecture, time structure, consequences of errors, technical robustness, clarity of information, and support logic play a decisive role in determining the conditions under which performance can be demonstrated at all. This shifts the question of the comparability of digital and analog exams toward institutional responsibility for the appropriateness of their respective exam conditions (Vindigni, 2025e).

In light of DIN EN ISO 9241-210, which explicitly describes human-centered design as an iterative and context-sensitive development process, this gives rise to the requirement that digital assessment environments be understood not as static implementations but as interactive systems that must be continuously evaluated (Deutsches Institut für Normung (DIN), 2020c; Geis & Tesch, 2023; Lestari, 2024; Vindigni, 2026c) . In this context, governance does not mean merely rule-setting or technical provision; it instead constitutes the institutional assurance that examination architectures, under real-world usage contexts, actually enable the conditions under which valid and fair performance assessment can take place (Vindigni, 2025e). The findings highlighted in the study regarding the importance of clear instructional logic, gradual activation, comprehensible communication structures, and learner-centered UX can therefore also be interpreted as governance-relevant insights: They identify criteria by which universities can assess whether digital examination formats are actually designed to be examination-appropriate or whether they generate additional, design-relevant requirements (Vindigni, 2025e). This means, this becomes particularly evident when digital exams are linked to implicit assumptions about technical availability, stable internet connections, uninterrupted concentration, or advanced technical proficiency. Such assumptions often seem trivial, but their cumulative effect is by no means neutral (Vindigni, 2025e). They can result in certain groups of students having to make systematically greater adjustments in order to meet the same exam requirements. From a governance perspective, this implies an obligation to reflect on exam architectures not only in terms of their functionality but also in terms of their implicit participation requirements. Digital exams are therefore not merely technical systems in the narrow sense but also institutional arrangements in which questions of equal opportunity, reasonableness, and the ability to participate are practically determined (Vindigni, 2025e).

Here, at this point, this becomes particularly evident when digital exams are linked to implicit assumptions about technical availability, stable internet connections, uninterrupted concentration, or advanced technical proficiency. Such assumptions often seem trivial, but their cumulative effect is by no means neutral. They can result in certain groups of students having to make systematically more significant adjustments to meet the same exam requirements, such as those with limited access to technology or those who are less familiar with digital platforms. Here and from a governance perspective, such reasoning implies an obligation to reflect on exam architectures not only in terms of their functionality but also in terms of their implicit participation requirements. Digital exams are therefore not merely technical systems in the narrow sense but also institutional arrangements in which questions of equal opportunity, reasonableness, and the ability to participate are practically determined (Vindigni, 2025e).

It follows, therefore, that the provision of alternative or compensatory examination formats should not be interpreted as a sign of inadequate digitalization; it should instead be recognized as an integral part of a responsible examination architecture. If digital examination environments generate increased interaction burdens or imply structural barriers to participation in certain contexts, it is appropriate from a governance perspective to provide hybrid or analog alternatives. It also becomes clear that the importance of context-sensitive design is evident in the study, which thus argues against viewing digital examinations as a universally superior standard format. Rather, it must be assumed that different examination formats may exhibit varying suitability profiles depending on the objectives, task structure, and usage context per se (Vindigni, 2025e).

Higher education institutions are thus faced with the task of implementing digital examinations technically and designing them as part of an overarching examination strategy in which validity, fairness, accessibility, and sensitivity to student workload are systematically integrated (Vindigni, 2025e). DIN EN ISO 9241 provides a suitable reference framework for these purposes because it explicitly links the design of digital systems to usage contexts, action goals, and human performance requirements. Applied to the examination context, this approach means that examination design must be understood not merely as a detailed pedagogical decision but as an integral part of institutional quality assurance. The governance perspective thus supplements the previously developed lines of argument with an organizational dimension: digital exams are not only a matter of individual use but also a matter of collective responsibility (Vindigni, 2025e).

Therefore, the analytical value of this case vignette lies in the fact that it translates the previously presented interaction and accessibility perspectives into an institutional framework (Vindigni, 2025e). It demonstrates that socio-technical examination architectures must be considered not only in terms of their immediate usability but also in terms of their structural embedding within organizational decision-making processes. Precisely because digital exams are increasingly becoming part of regular higher education practice, their design is not merely a question of technical innovation but a question of responsible implementation policy (Vindigni, 2025e). In this respect, this case vignette also confirms the basic assumption of this paper, namely that digital exams should be understood as interaction systems subject to normative evaluation, whose quality is determined not only by individual competence but also, and significantly, by institutional design decisions (Vindigni, 2025e).

CONCLUSION

This article has reconstructed digital exams in tertiary education as socio-technical interaction systems and has thereby challenged the reduction of performance differences to mere modality effects. The synthesis shows that outcomes in computer-based examinations are

conditioned by interaction quality, error consequences, context-of-use variability, monitoring regimes, and accessibility conditions. Against this background, DIN EN ISO 9241-11 is interpreted as a validity-relevant framework because effectiveness, efficiency, and satisfaction in a specific context of use can co-determine the construct interpretation of exam results (Deutsches Institut für Normung (DIN), 2018a).

DIN EN ISO 9241-110 provides the principle-based bridge for translating design decisions into fairness and outcome risks, while DIN EN ISO 9241-210 establishes the governance requirement to design, evaluate, and adjust digital exams through risk-appropriate human-centered processes (Deutsches Institut für Normung (DIN), 2020b; Deutsches Institut für Normung (DIN), 2020c).

The central implication is that universities should not treat digital examinations as neutral technical delivery formats or administrative efficiency measures. Rather, they must be governed as normatively assessable examination infrastructures. Where interaction load, error exposure, proctoring effects, and accessibility barriers remain unexamined, digitalization risks producing construct-irrelevant variance and new forms of exclusion. Conversely, systematically governed digital exams can contribute to a more valid, fair, accessible, and institutionally responsible assessment culture.

Acknowledgements

The author declares no actual or perceived conflicts of interest. This work received no external funding. The author confirms adherence to recognized ethical standards for social science publications; no new data involving human participants were collected for this study. Credit contribution: Conceptualization; Methodology; Investigation; Formal analysis; Writing—original draft; Writing—review & editing.

Declarations

This work was prepared in accordance with applicable principles of research integrity and transparency, including the German Research Foundation (DFG) Code of Conduct (Deutsche Forschungsgemeinschaft (DFG), 2025). Where relevant to the normative discussion of digitally mediated welfare-state contexts, established professional ethical guidance was consulted (e.g., National Association of Social Workers (National Association of Social Workers (NASW), 2021).

Availability of Data and Materials

No new primary data were collected for this study. The extraction and mapping scheme, the PRISMA-oriented documentation, and the supplementary matrices are provided as supplementary materials accompanying the article (PRISMA Executive, 2020).

Competing Interests

The author declares that there are no financial or other competing interests.

Funding

This study did not receive any earmarked financial support from public, commercial, or non-profit organizations.

Authors' Contributions

The author designed the study, developed the framework model, conducted the literature search and analysis (including PRISMA-oriented documentation), and wrote the manuscript (PRISMA Executive, 2020).

REFERENCES

Aini, Q., Yusup, M., Santoso, N. P. L., Ramdani, A. R., & Rahardja, U. (2021). Digitalization Online Exam Cards in the Era of Disruption 5.0 using the DevOps Method. *Journal of Educational Science and Technology (EST), 7*(1), 67–75. https://doi.org/10.26858/est.v7i1.18837

Al-Fraihat, D., Joy, M., Masa'deh, R., & Sinclair, J. (2020). Evaluating E-learning systems success: An empirical study. *Computers in Human Behavior, 102,* 67–86. https://doi.org/10.1016/j.chb.2019.08.004

Al-Qdah, M., & Ababneh, I. (2017). Comparing Online and Paper Exams: Performances and Perceptions of Saudi Students. *International Journal of Information and Education Technology, 7*(2), 106–109. https://doi.org/10.18178/ijiet.2017.7.2.850

Andreou, V., Peters, S., Eggermont, J., Wens, J., & Schoenmakers, B. (2021). Remote versus on-site proctored exam: comparing student results in a cross-sectional study. *BMC Medical Education, 21*(1). https://doi.org/10.1186/s12909-021-03068-x

Arksey, H., & O'Malley, L. (2005). Scoping studies: towards a methodological framework. *International Journal of Social Research Methodology, 8*(1), 19–32. https://doi.org/10.1080/1364557032000119616

Baragash, R. S., & Al-Samarraie, H. (2018). An empirical study of the impact of multiple modes of delivery on student learning in a blended course. *The Reference Librarian, 59*(3), 149–162. https://doi.org/10.1080/02763877.2018.1467295

Bashir, A., Bashir, S., Rana, K., Lambert, P., & Vernallis, A. (2021). Post-COVID-19 Adaptations; the Shifts Towards Online Learning, Hybrid Course Delivery and the Implications for Biosciences Courses in the Higher Education Setting. *Frontiers in Education, 6*. https://doi.org/10.3389/feduc.2021.711619

Browning, M. H. E. M., Larson, L. R., Sharaievska, I., Rigolon, A., McAnirlin, O., Mullenbach, L., Cloutier, S., Vu, T. M., Thomsen, J., Reigner, N., Metcalf, E. C., D'Antonio, A., Helbich, M., Bratman, G. N., & Alvarez, H. O. (2021). Psychological impacts from COVID-19 among university students: Risk factors across seven states in the United States. *PLOS ONE, 16*(1), e0245327. https://doi.org/10.1371/journal.pone.0245327

Bundesministerium für Digitales und

Staatsmodernisierung. (2025a). *Barrierefreie-Informationstechnik-Verordnung (BITV 2.0).*

Bundesministerium für Digitales und Staatsmodernisierung. (2025b). *Barrierefreiheitsstärkungsgesetz (BFSG).* https://www.barrierefreiheit-dienstekonsolidierung.bund.de/Webs/PB/DE/gesetze-und-richtlinien/barrierefreiheitsstaerkungsgesetz/barrierefreiheitsstaerkungsgesetz-node.html

Ceka, B., & O'Geen, A. J. (2019). Evaluating Student Performance on Computer-Based versus Handwritten Exams: Evidence from a Field Experiment in the Classroom. *PS - Political Science and Politics, 52*(4), 757–762. https://doi.org/10.1017/S104909651900091X

Chatzipetrou, P., Majkic, K., Samuelsson, J., & Åberg, O. (2023). Students' Acceptance Of Digital Exams: A Case Study In A Swedish University. *European Conference on E-Learning, 22*(1), 69–77. https://doi.org/10.34190/ecel.22.1.1602

Coghlan, S., Miller, T., & Paterson, J. (2021). Good Proctor or "Big Brother"? Ethics of Online Exam Supervision Technologies. *Philosophy & Technology, 34*(4), 1581–1606. https://doi.org/10.1007/s13347-021-00476-1

Cramp, J., Medlin, J. F., Lake, P., & Sharp, C. (2019). Lessons learned from implementing remotely invigilated online exams. In *Journal of University Teaching & Learning Practice* (Vol. 16, Number 1). https://ro.uow.edu.au/jutlpAvailableat:https://ro.uow.edu.au/jutlp/vol16/iss1/10

Daudt, H. M., van Mossel, C., & Scott, S. J. (2013). Enhancing the scoping study methodology: a large, inter-professional team's experience with Arksey and O'Malley's framework. *BMC Medical Research Methodology, 13*(1), 48. https://doi.org/10.1186/1471-2288-13-48

Davis, A. B., Rand, R., & Seay, R. (2016). Remote Proctoring: The Effect of Proctoring on Grades. In *Advances in Accounting Education: Teaching and Curriculum Innovations* (pp. 23–50). Emerald Group Publishing Limited. https://doi.org/10.1108/S1085-462220160000018002

Dendir, S., & Maxwell, R. S. (2020). Cheating in online courses: Evidence from online proctoring. *Computers in Human Behavior Reports, 2*. https://doi.org/10.1016/j.chbr.2020.100033

Deutsche Forschungsgemeinschaft (DFG). (2025, December 1). Gute wissenschaftliche Praxis. https://www.dfg.de/de/grundlagen-themen/grundlagen-und-prinzipien-der-foerderung/gwp

Deutsches Institut für Normung (DIN). (2009). DIN EN 614-1:2009-06: Sicherheit von Maschinen - Ergonomische Gestaltungsgrundsätze – Teil 1: Begriffe und allgemeine Leitsätze; Deutsche Fassung EN 614-1:2006+A1:2009.

Deutsches Institut für Normung (DIN). (2018a). DIN EN ISO 9241-11:2018-11, Ergonomie der Mensch-System-Interaktion – Teil 11: Gebrauchstauglichkeit: Begriffe und Konzepte (ISO 9241-11:2018); Deutsche Fassung EN ISO 9241-11:2018. DIN Media GmbH. https://doi.org/10.31030/2757945

Deutsches Institut für Normung (DIN). (2018b). DIN EN ISO 9241-306:2018-12, Ergonomie der Mensch-System-Interaktion – Teil 306: Vor-Ort-Bewertungsverfahren für elektronische optische Anzeigen (ISO_9241-306:2018); Deutsche Fassung EN ISO 9241-306:2018. DIN Media GmbH. https://doi.org/10.31030/2852270

Deutsches Institut für Normung (DIN). (2018c). ISO 9241-500:2018-11: Ergonomie der Mensch-System-Interaktion – Teil 500: Ergonomische Grundsätze zur Gestaltung und Beurteilung der Benutzungsumgebung.

Deutsches Institut für Normung (DIN). (2020a). DIN 33402-2:2020-12, Ergonomie – Körpermaße des Menschen – Teil 2: Werte. DIN Media GmbH. https://doi.org/10.31030/3170386

Deutsches Institut für Normung (DIN). (2020b). DIN EN ISO 9241-110:2020-10, Ergonomie der Mensch-System-Interaktion – Teil 110: Grundsätze der Dialoggestaltung (ISO 9241-110:2020); Deutsche Fassung EN ISO 9241-110:2020. Beuth Verlag GmbH. https://doi.org/10.31030/1464024

Deutsches Institut für Normung (DIN). (2020c). DIN EN ISO 9241-210:2020-03, Ergonomie der Mensch-System-Interaktion – Teil 210: Menschzentrierte Gestaltung interaktiver Systeme (ISO_9241-210:2019); Deutsche Fassung EN ISO9241-210:2019. DIN Media GmbH. https://doi.org/10.31030/3104744

Deutsches Institut für Normung (DIN). (2020d). ISO/TR 9241-514:2020-08: Ergonomie der Mensch-System-Interaktion – Teil 514: Leitfaden für die Anwendung anthropometrischer Daten in der. https://www.dinmedia.de/de/technische-regel/iso-tr-9241-514/328522258

Deutsches Institut für Normung (DIN). (2022a). DIN EN 301549:2022-06, Barrierefreiheitsanforderungen für IKT-Produkte und -Dienstleistungen; Englische Fassung EN_301549 V3.2.1 (2021-03); Text Deutsch, nur auf CD-ROM. DIN Media GmbH. https://doi.org/10.31030/3352891

Deutsches Institut für Normung (DIN). (2022b). DIN EN ISO 9241-20:2022-03, Ergonomie der Mensch-System-Interaktion – Teil 20: Ein ergonomischer Ansatz für die Barrierefreiheit innerhalb der ISO_9241-Reihe (ISO 9241-20:2021); Deutsche Fassung EN ISO 9241-20:2021. DIN Media GmbH. https://doi.org/10.31030/3325549

Deutsches Institut für Normung (DIN). (2022c). DIN EN ISO 15537:2022-09, Grundsätze für die Auswahl und den Einsatz von Prüfpersonen zur Prüfung anthropometrischer Aspekte von Industrieerzeugnissen und deren Gestaltung (ISO_15537:2022); Deutsche Fassung EN_ISO_15537:2022. DIN Media GmbH. https://doi.

org/10.31030/3332372

Eleanor, M. M., Ashley, K. E., & Morin, B. (2014). Comparing Student Performance on Computer-Based vs. Paper-Based Tests in a First-Year Engineering Course.

Elmer, T., Mepham, K., & Stadtfeld, C. (2020). Students under lockdown: Comparisons of students' social networks and mental health before and during the COVID-19 crisis in Switzerland. *PLOS ONE, 15*(7), e0236337. https://doi.org/10.1371/journal.pone.0236337

Erzurumlu, N. (2023). The Examination of Online and Paper-Pencil Test Scores of Students Engaged in Online Learning. *Asian Journal of Distance Education Taşkın, 18*(2). http://www.asianjde.com/

European Union. (2019). *Directive (EU) 2019/882 of the European Parliament and of the Council of 17 April 2019 on the accessibility requirements for products and services (Text with EEA relevance)*. https://eur-lex.europa.eu/eli/dir/2019/882/oj/eng

Geis, T., & Tesch, G. (2023). *Basiswissen Usability und User Experience: Aus- und Weiterbildung zum UXQB® Certified Professional for Usability and User Experience (CPUX) – Foundation Level (CPUX-F)*. Punkt.verlag.

Gibson, Y., & Kuhn, R. (2024). E-Proctoring Tools: Is It a Necessary Inconvenience. *Journal of Higher Education Theory and Practice, 24*(7). https://doi.org/10.33423/jhetp.v24i7.7088

Haddaway, N. R., Page, M. J., Pritchard, C. C., & McGuinness, L. A. (2022). PRISMA 2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis. Campbell Systematic Reviews, 18(2). https://doi.org/10.1002/cl2.1230

Hall, E. A., Roberts, M. B., Taylor, K. A., & Havrda, D. E. (2022). Changes in Academic Performance after Transitioning to Remote Proctoring: A Before-After Evaluation. *Pharmacy, 10*(4), 92. https://doi.org/10.3390/pharmacy10040092

Ita, M., Kecskemety, K., Ashley, K., & Morin, B. (2014). Comparing Student Performance on Computer-Based vs. Paper-Based Tests in a First-Year Engineering Course. *2014 ASEE Annual Conference & Exposition Proceedings,* 24.297.1-24.297.14. https://doi.org/10.18260/1-2--20188

Johnson, N., Veletsianos, G., & Seaman, J. (2020). U.S. Faculty and Administrators' Experiences and Approaches in the Early Weeks of the COVID-19 Pandemic. *Online Learning, 24*(2). https://doi.org/10.24059/olj.v24i2.2285

Jones, P., Tong, Y., Liu, J., Borglum, J., & Primoli, V. (2022). Score Comparability between Online Proctored and In-Person Credentialing Exams. *Journal of Educational Measurement, 59*(2), 180–207. https://doi.org/10.1111/jedm.12320

Khoshsima, H., Hosseini, M., & Toroujeni, S. M. H. (2017). Cross-Mode Comparability of Computer-Based Testing (CBT) Versus Paper-Pencil Based Testing (PPT): An Investigation of Testing Administration Mode among Iranian Intermediate EFL Learners. *English Language Teaching, 10*(2), 23. https://doi.org/10.5539/elt.v10n2p23

Komljenovic, J., & Williamson, B. (2024). Behind the platforms Safeguarding intellectual property rights and academic freedom in Higher Education. Higher Education. *Education Internationa, 69*.

Lestari, S. (2024). Does typing or handwriting exam responses make any difference? Evidence from the literature. *Research Matters: A Cambridge University Press & Assessment Publication, 38,* 66–81. https://doi.org/10.17863/CAM.111629

Mahyoob, M. (2020). Challenges of e-Learning during the COVID-19 Pandemic Experienced by EFL Learners. *Arab World English Journal, 11*(4), 351–362. https://doi.org/10.24093/awej/vol11no4.23

Malik, A. A., Hassan, M., Rizwan, M., Mushtaque, I., Lak, T. A., & Hussain, M. (2023). Impact of academic cheating and perceived online learning effectiveness on academic performance during the COVID-19 pandemic among Pakistani students. *Frontiers in Psychology, 14*. https://doi.org/10.3389/fpsyg.2023.1124095

Masterman, L., & Fresen, J. (n.d.). *Handwritten versus typed essay exams: a question of equivalence.* Retrieved www.digitaleducation.ox.ac.uk

Meulmeester, F. L., Dubois, E. A., Krommenhoek-van Es, C. (Tineke), de Jong, P. G. M., & Langers, A. M. J. (2021). Medical Students' Perspectives on Online Proctoring During Remote Digital Progress Test. *Medical Science Educator, 31*(6), 1773–1777. https://doi.org/10.1007/s40670-021-01420-w

Milone, A. S., Cortese, A. M., Balestrieri, R. L., & Pittenger, A. L. (2017). The impact of proctored online exams on the educational experience. *Currents in Pharmacy Teaching and Learning, 9*(1), 108–114. https://doi.org/10.1016/j.cptl.2016.08.037

Mogey, N., Paterson, J., Burk, J., & Purcell, M. (2010). Typing compared with handwriting for essay examinations at university: Letting the students choose. *ALT-J: Research in Learning Technology, 18*(1), 29–47. https://doi.org/10.1080/09687761003657580

Murphy, A. J., March-Dallas, S., & Tidwell, A. (2022). The Use of Remote Proctoring in Detecting Cheating: An Academic Dishonesty or Not Scenario. *Journal of Business and Social Science Review Issue, 3*(3), 1–11. https://doi.org/10.48150/jbssr.v3no3.2022.a1

National Association of Social Workers (NASW). (2021). *Code of Ethics.* https://www.socialworkers.org/About/Ethics/Code-of-Ethics/Code-of-Ethics-English

Newton, P. M. (2023). The validity of unproctored online exams is undermined by cheating. In *Proceedings of the National Academy of Sciences of the United States of America* (Vol. 120, Number 41). National Academy of Sciences. https://doi.org/10.1073/pnas.2312978120

Noorbehbahani, F., Mohammadi, A., & Aminazadeh, M. (2022). A systematic review of research on cheating in online exams from 2010 to 2021. *Education and Information Technologies, 27*(6), 8413–8460. https://doi.org/10.1007/s10639-022-10927-7

Oeding, J., Gunn, T., & Jones, A. (2024). The Impact of Remote Online Proctoring versus No Proctoring: A Study of Graduate Courses. *Online Journal of Distance Learning Administration, 27*(1), N/A.

Oeding, J., Gunn, T., & Seitz, J. (2024). The Mixed-Bag Impact of Online Proctoring Software in Undergraduate Courses. *Open Praxis, 16*(1), 82–93. https://doi.org/10.55982/openpraxis.16.1.585

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ, n71*. https://doi.org/10.1136/bmj.n71

PRISMA Executive. (2020). *PRISMA Statement*. https://www.prisma-statement.org/

Raaheim, A., Mathiassen, K., Moen, V., Lona, I., Gynnild, V., Bunæs, B. R., & Hasle, E. T. (2019). Digital assessment – how does it challenge local practices and national law? A Norwegian case study. *European Journal of Higher Education, 9*(2), 219–231. https://doi.org/10.1080/21568235.2018.1541420

Ravesloot, J. H., Donkers, J., Meiboom, A. A., Langers, A. M. J., jacobs, B., Tio, R., van der Vleuten, C., & Bremers, A. J. A. (2026). Computer Testing, Formative or Summative, and Proctoring: Does it Matter? Lessons Learned: From the Dutch Interuniversity Progress Test of Medicine During the Corona Pandemic. *Perspectives on Medical Education, 15*(1), 122–133. https://doi.org/10.5334/pme.1771

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research, 104,* 333–339. https://doi.org/10.1016/j.jbusres.2019.07.039

Torraco, R. J. (2005). Writing Integrative Literature Reviews: Guidelines and Examples. *Human Resource Development Review, 4*(3), 356–367. https://doi.org/10.1177/1534484305278283

Van der Weel, F. R., & Van der Meer, A. L. H. (2023). Handwriting but not typewriting leads to widespread brain connectivity: a high-density EEG study with implications for the classroom. *Frontiers in Psychology, 14*. https://doi.org/10.3389/fpsyg.2023.1219945

van Halem, N., van Klaveren, C., & Cornelisz, I. (2021). The effects of implementation barriers in virtually proctored examination: A randomised field experiment in Dutch higher education. *Higher Education Quarterly, 75*(2), 333–347. https://doi.org/10.1111/hequ.12275

Vindigni, G. (2025a). Between Access Rights and Copyright: The European Accessibility Act as a Regulatory Framework for AI-Generated Content. *European Journal of Management, Economics and Business, 2*(1), 31–56. https://doi.org/10.59324/ejmeb.2025.2(1).03

Vindigni, G. (2025b). Data-Driven Disparities: How AI Applications in Education May Perpetuate or Mitigate Inequality. *EJSMT, 1*(4), 4–54. https://doi.org/10.59324/ejsmt.2025.1(4).02

Vindigni, G. (2025c). Digital Teaching-Learning Processes in Tertiary Education: An Empirical Analysis of Zoom Usage, Cognitive Control, and Psychological Stress from the Normative Perspective of DIN EN ISO 9241. *European Journal of Contemporary Education and E-Learning, 3*(6), 101–135. https://doi.org/10.59324/ejceel.2025.3(6).09

Vindigni, G. (2025d). Digital Transformation as a Challenge for Social Sustainability: Participation, Fairness, and Inclusion in Child and Youth Services in the Age of AI-Based Systems. *European Journal of Innovative Studies and Sustainability, 2*(1), 1–53. https://doi.org/10.59324/ejiss.2026.2(1).01

Vindigni, G. (2025e). Gender Bias and Cultural Misrepresentation in AI: A Critical Inquiry into Cross-Cultural Communication and Algorithmic Design. *European Journal of Applied Science, Engineering and Technology, 3*(3), 51–72. https://doi.org/10.59324/ejaset.2025.3(3).06

Vindigni, G. (2025f). The Paradox of Decentralization: NFTs, Platform Capitalism and the Tokenized Knowledge Economy. *Scientia. Technology, Science and Society, 2*(11), 132–177. https://doi.org/10.59324/stss.2025.2(11).13

Vindigni, G. (2026a). Democratization through Tokenization? Digital Educational Credentials between Decentralization and Re-Intermediation – A Critical Governance Analysis. *Journal of Artificial Intelligence and Technological Development, 2*(2), 46–73. https://doi.org/10.59324/jaitd.2026.2(2).04

Vindigni, G. (2026b). Digital by Default and Digital Exclusion: Accessibility Regimes, Access Conditionality, and the Right to Analog Participation – An International Systematic Research Overview. *European Journal of Arts, Humanities and Social Sciences, 3*(2), 28–90. https://doi.org/10.59324/ejahss.2026.3(2).0

Vindigni, G. (2026c). *UX Research und User Research: Theorie, Methoden und Anwendungsfelder der Nutzerforschung*. Springer Vieweg. https://doi.org/10.1007/978-3-662-72235-0

Vindigni, G. (2026d). Beyond Compliance: Future Skills 2030, Digital Accessibility and Democratic Participation in Higher Education – A Scoping Review. *European Journal of Contemporary Education and E-Learning, 4*(4), 56-137. https://10.59324/ejceel.2026.4(4).05.

W3C Web Accessibility Initiative. (2025). *WCAG 2 Overview*. https://www.w3.org/WAI/standards-guidelines/wcag/

Wacker, J. G. (1998). A definition of theory: research guidelines for different theory-building research methods in operations management. Journal of *Operations Management, 16*(4), 361–385. https://doi.org/10.1016/S0272-6963(98)00019-9

Wagner, I., Loesche, P., & Bißantz, S. (2022). Low-stakes performance testing in Germany by the VERA assessment: analysis of the mode effects between computer-based testing and paper-pencil testing. *European Journal of Psychology of Education, 37*(2), 531–549. https://doi.org/10.1007/s10212-021-00532-6

Wang, S., Jiao, H., Young, M. J., Brooks, T., & Olson, J. (2008). Comparability of computer-based and paper-and-pencil testing in K-12 reading assessments: A meta-analysis of testing mode effects. *Educational and Psychological Measurement, 68*(1), 5–24. https://doi.org/10.1177/0013164407305592

Weirich, S., Sachse, K. A., Henschel, S., & Schnitzler, C. (2024). Comparing Test-Taking Effort Between Paper-Based and Computer-Based Tests. *Applied Psychological Measurement, 48*(1–2), 3–17. https://doi.org/10.1177/01466216241227535

Whittemore, R., & Knafl, K. (2005). The integrative review: updated methodology. *Journal of Advanced Nursing, 52*(5), 546–553. https://doi.org/10.1111/j.1365-2648.2005.03621.x

Woldeab, D., & Brothen, T. (2024). Exploring the Impact: How Online Exam Proctoring Reduces Cheating and Enhances Course Legitimacy. *International Journal of E-Learning & Distance Education / Revue Internationale Du e-Learning et La Formation à Distance, 39*(1). https://doi.org/10.55667/10.55667/ijede.2024.v39.i1.1327

Yanfei, M. (2021). Online and Offline Mixed Intelligent Teaching Assistant Mode of English Based on Mobile Information System. *Mobile Information Systems, 2021,* 1–6. https://doi.org/10.1155/2021/7074629

Supplementary Materials

Supplement A: Search and Selection Strategy (PRISMA-oriented), Including Search Areas and Search Strings

A.1 Search Period and Language Criteria

The search and selection period were set from January 2008 to February 2026 in order to capture (i) the phase of consolidation of computer-based exam formats in tertiary education, (ii) the development of remote proctoring and integrity regimes, and (iii) the increasing relevance of usability, accessibility, and governance aspects in the context of digital examinations in their longitudinal dynamics. Primarily English- and German-language publications were considered; additional languages were only included if they addressed specific regime constellations or normatively relevant implementation contexts and thus contributed substantially to the scope of the synthesis.

A.2 Information Sources and Search Areas

To ensure an interdisciplinary evidence base, bibliometric indices, subject-specific databases, publishing platforms, preprint/open science repositories, and normative and regulatory sources were combined. The search areas included in particular Scopus, Web of Science Core Collection, ERIC, PsycINFO, PubMed/MEDLINE, CINAHL, IEEE Xplore, ACM Digital Library, INSPEC, Education Source, EBSCO Academic Search Complete, SocINDEX, ProQuest Central, ProQuest Dissertations & Theses, ScienceDirect, SpringerLink, Wiley Online Library, Taylor & Francis Online, SAGE Journals, Emerald Insight, JSTOR, arXiv, SSRN, DOAJ, BASE, OpenAIRE, CORE, and Google Scholar to cover gray literature (reports, working papers, and preprints). In addition, normative/regulatory sources were systematically reviewed (W3C/WAI for WCAG documents, ETSI for EN 301 549 materials, and EUR-Lex for EAA/BFSG-related legal sources) where they were necessary for contextualizing the accessibility and governance dimensions.

A.3 Search Logic and Thematic Clusters

The search strings were operationalized as combined Boolean queries along three analytical complexes:

(i) performance and comparison perspective (CBT vs. paper/handwriting; mode/format effects; comparability),

(ii) integrity and monitoring regime (proctoring, cheating, integrity, surveillance),

(iii) socio-technical interaction conditions (usability/HCI/ergonomics/ISO 9241; consequences of errors; context of use; accessibility/compliance).

A.4 Example Search Strings (title/abstract/keywords; adapted according to database syntax)

Cluster 1: CBT vs. paper/handwriting/comparability

• ("computer-based test*" OR "computer-based exam*" OR CBT OR "online exam*" OR "digital exam*") AND (paper OR "paper-based" OR "handwritten" OR handwriting OR "pencil-and-paper" OR PBT) AND (comparab* OR equivalen* OR "mode effect*" OR "format effect*" OR performance OR score*) AND (higher education OR universit* OR college* OR "tertiary education")

Cluster 2: Writing Mode Mechanisms (handwriting vs. typing)

• (handwriting OR "hand-written" OR "note taking" OR typing OR keyboard* OR "writing mode*") AND (mechanism* OR cognitive OR "text production" OR revision OR planning OR "working memory") AND (exam* OR assessment OR test*) AND (university OR "higher education" OR "tertiary education")

Cluster 3: Proctoring/integrity

• ("remote proctor*" OR "online proctor*" OR "exam proctor*" OR "test security" OR "academic integrity" OR cheating) AND (exam* OR assessment OR test*) AND ("higher education" OR universit* OR college* OR "tertiary education") AND (grade* OR performance OR outcome*)

Cluster 4: Usability/HCI/ISO 9241/Interaction quality/Consequences of errors

• (usability OR "human-computer interaction" OR HCI OR ergonom* OR "ISO 9241" OR "interaction principle*" OR "error tolerance" OR "error consequence*" OR "context of use") AND (exam* OR assessment OR "testing system" OR "assessment platform") AND ("higher education" OR universit* OR "tertiary education")

Cluster 5: Accessibility/Compliance (WCAG/BITV/EN 301 549/EAA/BFSG)

• (accessibilit* OR WCAG OR "EN 301 549" OR BITV OR "European Accessibility Act" OR EAA OR BFSG) AND (exam* OR assessment OR "learning management system" OR LMS OR "assessment platform") AND ("higher education" OR universit* OR "distance education")

A.5 Screening, Deduplication, and Reasons for Exclusion

After merging the hits into a common reference pool, deduplication was performed (automatically and manually). Screening was carried out in two stages (title/abstract, then full text). Reasons for exclusion were categorized and documented for PRISMA (see Supplement D).

Supplement B: Extraction Scheme and Codebook (ISO 9241 mapping)

B.1 Purpose and Structure

The extraction scheme serves to ensure the intersubjective transparency of evidence triangulation by collecting a uniform set of variables for each source that makes the findings a function of (i) the document, (ii) the examination constellation, and (iii) the assignment to ISO-9241 anchors and effect path types traceable.

B.2 Extraction Variables (codebook)

Bibliographic Metadata

• Author(s), year, title, source/publication type (peer-reviewed; report; dissertation; preprint)

• Country/regime context (if relevant), field of study/discipline, setting (campus/remote/hybrid)

Evidence mode and methodological type

• Comparative study (CBT vs. PBT/handwriting) / field experiment (crossover) / quasi-experiment / observational study / review / mechanism study / proctoring study / standard/policy document

• Sample (N), population (tertiary), task format (MC/essay/short answer/mixed)

Examination system and context of use (ISO 9241-11; supplemented by environment/display/anthropometry where relevant)

• Device policy (BYOD/managed), network/environmental conditions, support/support paths

• Accessibility/assistive tech conditions, room/remote constraints

• Error handling regime (timeouts, termination rules, sanctions, recovery options)

Interaction principles/UI properties (ISO 9241-110)

• Task appropriateness, self-descriptiveness, conformity to expectations, controllability, error tolerance, customizability, learning supportiveness

• Operationalizable indicators (e.g., editing effort, navigation load, error events, recovery rate)

Proctoring/integrity as a contextual intervention

• Proctoring type (live/recorded/AI), intervention logic (lockdown, monitoring), false positives/disruption events (if reported)

Outcomes/result dimensions

• Performance/score, processing time, dropout rate, error rate

• Acceptance/preference, stress/strain indicators (if reported)

• Validity/fairness notes (explicit/implicit)

Assignment

• ISO anchor: 9241-11 / 9241-110 / 9241-210 (supplementary standards as context anchors, if applicable)

• Effect path type: Context variance / Interaction load / Error consequences / Monitoring regime / Accessibility conditions (others, if applicable)

Supplement C: Complete mapping and path matrix

C.1 Content and reading logic

For each source included, the complete mapping matrix documents the extraction results and the assignment to ISO 9241 categories and effect pathway types. It serves as an audit trail for the synthesis and enables the traceability of the assignments summarized in the results section and visualized in Figure 2.

C.2 Column structure (recommended)

• Study ID
• Evidence mode / study type
• Setting (tertiary; campus/remote)
• Exam format (CBT/PBT; response type)
• Key finding (short)
• Outcome dimension (performance/time/errors/acceptance/strain)
• Path type(s) (context; interaction; error consequences; monitoring; accessibility)
• ISO anchor(s) (9241-11/-110/-210)
• Operationalization hint (minimal indicator)
• Notes (boundary conditions; limitations)

Supplement D: PRISMA-compliant Exclusion Categories (title/abstract and full text)

D.1 Exclusion categories Title/abstract screening

• Out of scope (no examination/assessment context)
• Not tertiary or no justified transferability
• Purely technical system tests without reference to examination outcomes/validity/fairness
• Insufficient conceptual or methodological substance (e.g., editorials without a reliable basis)

• No relevance to the analysis complexes (comparability / socio-technical interaction / proctoring integrity / accessibility compliance)

D.2 Exclusion categories Full-text screening

• No comparable examination mode or no evaluable outcome dimension

• Insufficient reporting transparency (e.g., missing information on setting/population/instrument)

• No tertiary context and no justified transferability

• Non-empirical opinion formats without methodological foundation

• Redundant publications/multiple versions without added value