

A Motivation-Aware Diagnostic Instrument for Personalized and Gamified Computing Education

Sergio Antônio Andrade de Freitas, Daniel Rodrigues da Rocha,
Davi Rodrigues da Rocha, Mylena Angélica Silva Farias

¹Centro de Estudos, Desenvolvimento e Inovação em Software (CEDIS)
University of Brasília (UnB) – Brasília, DF – Brazil

Corresponding author: sergiofreitas@unb.br

Abstract. *Semester-start evidence can inform instructional adaptation before disengagement or low performance becomes visible. This paper presents a preliminary empirical assessment of a motivation-aware diagnostic instrument used in three undergraduate computing courses, with 245 valid responses: 81 from Human–Computer Interaction, 133 from Compilers, and 31 from Introduction to Gamification. The instrument combines course-specific self-assessments of prior knowledge with common blocks on instructional perceptions, challenge, autonomy, performance expectations, social connection, curiosity, and course motivation. The study positions the instrument as a screening and planning aid rather than a validated psychometric scale or an automatic intervention engine. The combined common-item subset showed preliminary internal consistency ($\alpha = 0.880$). Small-sample reliability estimates were treated as exploratory, and selected correlations were reported with confidence intervals. The main contribution is an operational protocol that links diagnostic patterns to universal, targeted, and short-cycle instructional responses, including feasible options for heterogeneous classes and constrained computing curricula.*

Keywords *Computing education, Diagnostic assessment, Gamification, Motivation, Learning analytics, Personalized learning*

1. Introduction

Diagnostic assessment, formative assessment, and adaptive learning already constitute established research traditions. They treat evidence about learners not merely as information to be collected, but as part of a cycle that connects observations, interpretations, and instructional action [Black and Wiliam 1998, Pellegrino et al. 2001, Mislevy et al. 2003, Brusilovsky 2001, Bernacki et al. 2021]. The problem addressed here is therefore not the absence of diagnostic instruments in education. It is the narrower operational problem of combining semester-start evidence about course-specific prior knowledge and motivation with a transparent, low-overhead decision process for undergraduate computing courses.

This problem is relevant because personalization is often implemented only after grades, attendance, or platform traces reveal disengagement. Earlier evidence may allow instructors to adjust examples, pacing, feedback, grouping, and optional task paths before difficulties become entrenched. However, adaptation cannot be reduced to assigning one intervention to one score. Computing classes commonly contain students with different levels of prior knowledge, confidence, autonomy, and perceived relevance, while instructors face fixed learning outcomes, large enrollments, limited time, and mandatory syllabi. A

useful diagnostic instrument must therefore support decisions for coexisting patterns within the same class rather than assume a homogeneous cohort.

Motivation-aware design adds a further requirement. Personalized gamification can affect motivation, but its effects depend on context and learner characteristics [Rodrigues et al. 2024, Lavoué et al. 2019]. Prior work from the same research program examined motivational profiling and the use of gamification with learning indicators [de Freitas et al. 2024a, de Freitas et al. 2024b]. The present study contributes a complementary baseline layer: a questionnaire administered at the start of the semester and an explicit protocol for translating its outputs into feasible instructional hypotheses. The instrument is not positioned as a substitute for performance-based diagnostic tasks, teacher judgment, or longitudinal learning analytics.

The study is guided by three research questions:

- RQ1.** What preliminary evidence of internal consistency and interpretability does the questionnaire provide for semester-start diagnostic use in computing courses?
- RQ2.** Which exploratory patterns of self-assessed prior knowledge, instructional perception, expectations, autonomy, curiosity, and course motivation emerge across the three courses?
- RQ3.** How can these indicators inform feasible personalization and gamification decisions when heterogeneous student patterns coexist under curricular and resource constraints?

This paper makes four contributions. First, it presents a reusable instrument that combines course-specific prior-knowledge items with common motivational and instructional dimensions. Second, it reports descriptive, internal-consistency, and correlational evidence from 245 responses, while explicitly acknowledging limitations related to small subsamples and measurement design. Third, it brings together perspectives from diagnostic assessment, adaptive learning, motivation, and differentiated instruction. Fourth, it introduces a tiered decision protocol comprising universal support, targeted options, and short-cycle follow-up. This protocol preserves shared course objectives while allowing feasible adjustments to support, feedback, grouping, and task pathways.

2. Background and related work

2.1. Diagnostic assessment and learner adaptation

A diagnostic process is useful when evidence changes what happens next. Formative-assessment research emphasizes feedback loops in which evidence is elicited, interpreted, and used to adapt teaching and learning [Black and Wiliam 1998]. Evidence-centered assessment similarly distinguishes observations, claims about learners, and the tasks or situations that justify those claims [Pellegrino et al. 2001, Mislevy et al. 2003]. This distinction is important here: a questionnaire response is an observation that supports an instructional hypothesis, not a definitive classification of a student.

Prior knowledge is central to this reasoning because the effectiveness of instructional support depends on what learners already know. The expertise-reversal effect, for example, shows that guidance useful to novices may become redundant or counterproductive for more knowledgeable learners [Kalyuga 2007]. Computing education also has

an established diagnostic-assessment tradition. A systematic review cataloged 33 computing concept inventories, 12 of which had been validated, for identifying conceptual understanding and misconceptions [Ali et al. 2023]. Consequently, a semester-start screen can support differentiated scaffolding, but self-assessed familiarity should be distinguished from objective diagnostic evidence and complemented by brief performance-based checks whenever decisions carry substantial consequences.

Adaptive-learning research has long modeled learner characteristics to select content, navigation, feedback, or support [Brusilovsky 2001]. A recent systematic review of personalized learning likewise shows that effective personalization depends on how learner information is collected, interpreted, and connected to interventions, not on data collection alone [Bernacki et al. 2021]. A computing-education scoping review identified 29 adaptive-learning studies, concentrated in programming and databases, and reported that prior knowledge is commonly represented in student models while adaptation often targets learning resources [Barbosa et al. 2024]. The present instrument occupies a place in this landscape: it provides an interpretable baseline for instructor-led adaptation rather than a fully automated adaptive system.

2.2. Motivation, gamification, and instructional design

Gamification is commonly defined as the use of game design elements in non-game contexts [Deterding et al. 2011]. Educational applications can structure progress, reinforce feedback cycles, and foster participation, but effects vary with context, target population, and design quality [Hamari et al. 2014, Caponetto et al. 2014, Dehghanzadeh et al. 2024, Zeybek and Saygı 2024]. Adaptive gamification therefore requires evidence about learners and should avoid assuming that the same mechanics benefit everyone [Lavoué et al. 2019, Rodrigues et al. 2024].

Self-Determination Theory (SDT) emphasizes autonomy, competence, and relatedness as basic psychological needs associated with the quality of motivation [Ryan and Deci 2000]. The Intrinsic Motivation Inventory (IMI) operationalizes related dimensions such as interest/enjoyment, perceived competence, effort/importance, pressure/tension, value/usefulness, perceived choice, and relatedness [Ryan et al. 1983]. Keller's ARCS model complements this perspective through Attention, Relevance, Confidence, and Satisfaction [Keller 2010]. Reviews and meta-analyses indicate that ARCS can inform both instructional design and motivational measurement [Fang et al. 2024, Dinçer 2020, Göksu and Bolat 2021]. These theories support interpretations such as using guided practice for low confidence, authentic problems for low relevance, or optional challenge paths for high autonomy and curiosity. They do not, by themselves, prescribe a unique intervention.

2.3. Heterogeneous classes and feasible differentiation

Differentiated instruction addresses variation within a shared classroom by modifying support, process, grouping, or products while maintaining common learning goals [Tomlinson 2017]. Evidence from secondary-education research suggests potential benefits, but also shows that implementation quality and contextual constraints matter [Smale-Jacobse et al. 2019]. In large university classes, full individual learning paths may be infeasible. A more realistic design is a common instructional core combined with a small number of targeted supports and choices. This principle informs the decision protocol proposed in Section 4.3:

Table 1. Participant profile by course. The final column reports students taking the course for the first time and students with prior experience with the same instructor.

Course	<i>n</i>	Age mean	Age SD	Female	Male	Unreported	First time / prior instructor
HCI	81	20.94	1.52	8	72	1	79 / 27
Compilers	133	21.87	2.41	18	115	0	121 / 54
Intro to Gamification	31	22.29	3.06	5	26	0	30 / 19
Combined	245	21.62	2.33	31	213	1	230 / 100

diagnostic patterns are used to prioritize low-overhead responses and are revisited through short-cycle evidence rather than converted into permanent labels.

3. Research design and methodology

3.1. Study design and context

The study was a cross-sectional, multi-course investigation conducted at the beginning of the second 2025 academic term at a Brazilian public university. The questionnaire was administered during the second class meeting of three undergraduate computing courses: Human–Computer Interaction (HCI), Compilers, and Introduction to Gamification. Its immediate purpose was instructional diagnosis: to characterize initial conditions and inform planning, gamification choices, and subsequent monitoring.

The data set contained 245 valid responses: 81 from HCI, 133 from Compilers, and 31 from Introduction to Gamification. All analyzed records corresponded to students who accepted the consent statement. The combined sample had a mean age of 21.62 years ($SD = 2.33$), with 213 male respondents, 31 female respondents, and one unreported gender value. Most students were taking the course for the first time (230/245). A conditional branch asked whether students had previously taken another course with the same instructor; this branch explains smaller effective samples in analyses involving prior-instructor perception. Table 1 summarizes the participant profile.

3.2. Diagnostic workflow and data preparation

Figure 1 summarizes the instructor-in-the-loop process that connects semester-start evidence with revisable instructional action. Responses are summarized as block distributions and provisional subgroup patterns, interpreted in light of theory and course constraints, and translated into universal support, targeted options, or both. Short-cycle evidence is then used to retain, revise, or discontinue the selected responses.

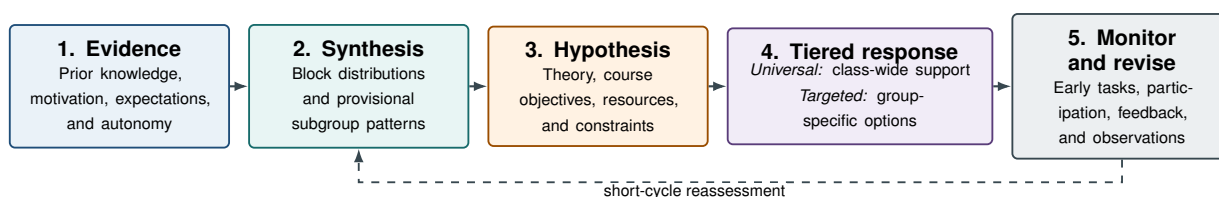


Figure 1. Instructor-in-the-loop workflow connecting semester-start evidence to provisional interpretation.

Data preparation followed four steps. First, the original questionnaire labels were mapped to shorter and analytically tractable identifiers. Second, Likert-type response categories were converted into numeric values while preserving their original ordering and scale ranges. Third, negatively worded items were reverse coded so that higher values consistently represented more positive perceptions, expectations, or motivational conditions. Fourth, conceptually related items were grouped into exploratory composites by averaging the available valid responses within each block.

The historical questionnaires contained both 5-point and 4-point sections. The course-motivation block was inherited from an earlier instrument that used four response categories, whereas the remaining Likert-type blocks used five categories. This heterogeneity reflects the provenance of the instruments rather than an analytic decision made after examining the data.

No normalization was applied across the two scale formats. A linear range transformation would align numerical endpoints but would not establish measurement equivalence between constructs or response categories. It could also obscure the interpretation of the original scales. The 4-point course-motivation block was therefore analyzed and reported separately from the 5-point blocks and was not combined with them to produce an overall score.

Cross-course comparisons were conducted only within the same composite, ensuring that each comparison involved a common response scale and item structure. Pearson correlations are mathematically invariant under positive linear rescaling; however, their substantive interpretation still depends on construct validity, response distributions, and effective sample size. Descriptive means were consequently retained on their original metrics to preserve interpretability.

Conditional branching and missing responses produced different effective sample sizes across analyses. These values are reported for each result. Descriptive composites were calculated from the available valid responses within each block, whereas internal-consistency estimates were based on complete cases for the corresponding item set.

3.3. Ethical procedures

Participation was voluntary and preceded by an informed-consent statement. Analyses used anonymized records and report only aggregate results.

4. Instrument structure

4.1. Course-specific and common blocks

The questionnaire combines course-specific and common components. The prior-knowledge block was customized because the three courses emphasize different technical concepts. HCI items covered usability, accessibility, formative evaluation, UX metrics, Likert scales, Gestalt principles, and prototyping tools [Rogers et al. 2023]. Compilers items covered automata, grammars, lexical and syntactic analysis, semantic analysis, intermediate code, optimization, and compiler organization [Aho et al. 2006]. Introduction to Gamification items covered intrinsic and extrinsic motivation, motivational profiles, ethics, Octalysis [Chou 2015], SDT [Ryan and Deci 2000], MDA [Hunicke et al. 2004], the 6D framework [Werbach and Hunter 2012], responsible design, and personalization.

Table 2. Questionnaire structure. Prior knowledge is self-assessed; the other blocks capture perceptions or expectations at semester start.

Block	Scope	Items	Diagnostic purpose
Demographics and preferences	Common	9	Age, gender, academic status, and preferences for practical, expository, and group formats.
Prior knowledge	Course-specific	HCI: 21; Compilers: 15; Intro: 24	Self-assessed familiarity with course concepts, techniques, and tools; not an objective mastery test.
Current instructor perception	Common	2	Perceived ability to generate enthusiasm and use varied teaching techniques.
Prior instructor perception	Conditional common	6	Perceptions of examples, respect, feedback, critical thinking, professional relevance, and mastery in previous courses with the same instructor.
Expectations and challenge	Common	9	Expected challenge, relevance, effort, monotony, interest, and anticipated pleasure.
Motivation and autonomy	Common	9	Autonomy, perceived choice, expected competence gains, and effort intentions.
Performance, social connection, and curiosity	Common	11	Expected performance, relatedness, anxiety, pressure, nervousness, curiosity, and willingness to retake the course.
Course motivation and engagement	Common	15	Course-level motivation and engagement on a separate 4-point scale.

Table 3. Theory-to-decision mapping. Each action is an instructional hypothesis to be checked against course constraints and later evidence.

Construct	Source	Diagnostic evidence	Candidate instructional action
Competence and confidence	SDT, IMI, ARCS	Low perceived prior knowledge; low expected success; high challenge	Verify with a brief task; use worked examples, early formative feedback, and gradual task complexity.
Autonomy and choice	SDT, IMI	High voluntary participation; perceived choice; willingness to engage	Offer bounded project choices, optional challenge paths, and flexible activity formats.
Relevance and value	ARCS, IMI	Low perceived utility or weak professional connection	Use authentic problems, professional cases, and explicit links to computing practice.
Attention and curiosity	ARCS	High curiosity or interest	Use exploratory tasks, open-ended questions, demonstrations, or narrative hooks.
Relatedness and trust	SDT	Respect, feedback, belonging, or instructor-trust indicators	Preserve instructor presence, collaborative activities, inclusive grouping, and timely feedback.
Pressure and anxiety	IMI	Nervousness, pressure, or negative affect	Reduce early stakes, clarify success criteria, and provide private low-stakes rehearsal before grading.

The common blocks support cross-course comparison of class preferences, instructor perception, expectations and challenge, motivation and autonomy, performance and social connection, engagement and curiosity, and course motivation. Table 2 summarizes the structure. Minor differences among historical forms explain why complete-case analyses and combined common-item analyses use different item counts. The prior-knowledge responses represent perceived familiarity, not demonstrated mastery.

4.2. Theory-to-decision mapping

Table 3 provides a transparent bridge from diagnostic evidence to instructional hypotheses. The actions are options to be evaluated in context, not automatic prescriptions. For example, low self-assessed knowledge and low expected success justify checking whether early scaffolding is needed; high autonomy and curiosity justify considering choices or optional challenge paths. In gamified courses, this logic prioritizes mechanics that support competence, autonomy, relatedness, and meaningful progress over isolated points or badges.

4.3. Decision protocol for heterogeneous classes

Because multiple patterns can coexist, the unit of action is not necessarily the entire cohort or an individual student. The protocol uses three levels. First, a *universal core* addresses needs shared by a substantial portion of the class without changing mandatory outcomes. Second, *targeted options* provide low-overhead variation through flexible groups, optional

Table 4. Low-overhead responses when heterogeneous patterns coexist in one class. Mandatory outcomes remain common to all students.

Observed pattern	Universal core action	Targeted option	Short-cycle check
Broad low prior knowledge and high challenge	Brief diagnostic task, concept review, worked example, and clear success criteria	Guided practice sequence and optional support session	Exit ticket or low-stakes quiz after one or two topics
Mixed prior knowledge within the same class	Common minimum task and explicit prerequisite map	Temporary flexible groups; scaffolded and extension versions of the same objective	Compare completion, errors, and help requests by task path
High autonomy or curiosity in a subgroup	Explain common outcomes and provide meaningful rationale	Bounded choice of project domain, optional challenge, or exploratory task	Track uptake, completion, and reflective feedback
Low relevance or value	Connect the topic to an authentic computing problem	Choice among professional cases with equivalent objectives	One-item relevance check plus task participation
High pressure, anxiety, or low belonging	Normalize practice, lower early stakes, and make expectations explicit	Private rehearsal, peer support, or instructor follow-up; avoid public ranking	Monitor non-submission, help seeking, and affective recheck

support, challenge extensions, or bounded choices. Third, *focused follow-up* is reserved for a small number of students whose combined indicators and early performance suggest persistent difficulty. Table 4 operationalizes this logic. Thresholds should be based on the observed distribution, course requirements, and instructor judgment rather than treated as validated clinical or psychometric cutoffs.

5. Data analysis

Descriptive analyses summarized each composite by course. Because conditional branching and partial missingness produced different effective samples, every table reports the relevant n . No total score combining the 4-point and 5-point blocks was computed.

Exploratory one-way ANOVAs screened for cross-course mean differences within seven composites. The analysis reports F with degrees of freedom, exact p values, and omega squared (ω^2) as an effect-size estimate. Unequal group sizes and Likert-derived composites require conservative interpretation; no post-hoc pairwise claims are made. Because seven omnibus tests were inspected, results are also discussed against a conservative Bonferroni reference of $\alpha = 0.05/7 = 0.0071$. This reference is a sensitivity check rather than a preregistered confirmatory decision rule.

Internal consistency was assessed with Cronbach's alpha [Tavakol and Dennick 2011]. Two kinds of item sets were retained: broad reconstructed non-course sets within HCI and Compilers, and a 39-item common subset pooled across courses. The broad sets contain more items than complete cases ($k > n$), so their coefficients are reported only as descriptive audits and not as evidence of unidimensionality. The 15-item course-motivation block was also examined separately by course. A previously generated stabilized coefficient for the 64-item Introduction to Gamification non-course set was not used as primary evidence because the retained analysis record did not document the stabilization procedure sufficiently for replication. Alpha is not treated as proof of factorial, convergent, or predictive validity; future studies should add McDonald's omega and ordinal reliability estimates [Malkewitz et al. 2023].

Pearson correlations were used as exploratory association screens. To reduce overinterpretation, Table 7 omits exact $r = 1.000$ pairs from very small samples and reports selected non-perfect associations with Fisher-transformed 95% confidence intervals. No multiplicity adjustment was applied to these selected correlations, so they are interpreted

Table 5. Composite means on their original scales. Cells report mean (SD; effective n). The 4-point course-motivation block is not directly compared with 5-point block means.

Composite	Scale	HCI	Compilers	Intro to Gamification
Prior knowledge	1–5	1.66 (0.56; 81)	1.60 (0.65; 133)	2.03 (0.79; 31)
Current instructor perception	1–5	4.73 (0.65; 79)	4.55 (0.63; 133)	4.73 (0.47; 30)
Expectations and challenge	1–5	4.48 (0.42; 27)	4.24 (0.48; 88)	4.47 (0.30; 19)
Motivation and autonomy	1–5	4.32 (0.49; 27)	4.19 (0.48; 88)	4.56 (0.28; 19)
Performance and social connection	1–5	3.71 (0.70; 27)	3.60 (0.53; 88)	3.68 (0.56; 19)
Engagement and curiosity	1–5	3.87 (0.89; 27)	3.64 (0.78; 88)	4.05 (0.60; 19)
Course motivation	1–4	2.79 (0.61; 81)	2.83 (0.62; 133)	2.73 (0.51; 31)

as candidate relationships for replication rather than confirmed construct links. Composite-level correlations were analyzed in the same exploratory manner. Causal inference is outside the scope of the cross-sectional design.

6. Results

6.1. Sample profile and descriptive tendencies

The combined sample contained 245 responses, with the largest group in Compilers ($n = 133$), followed by HCI ($n = 81$) and Introduction to Gamification ($n = 31$). The number of students who had previously taken another course with the same instructor was 54 in Compilers, 27 in HCI, and 19 in Introduction to Gamification. These values explain the smaller effective samples in conditional blocks.

Table 5 presents means on the original scales, with standard deviation and effective n in parentheses. Self-assessed prior knowledge was low in all three courses, particularly HCI and Compilers. Current-instructor perception was high across courses. Expectations/challenge and motivation/autonomy were above the scale midpoint, whereas course motivation on its separate 4-point scale was moderate. These descriptions concern course-level distributions and do not establish homogeneous class profiles.

The exploratory ANOVAs indicated differences in prior knowledge, $F(2, 242) = 6.544$, $p = 0.0040$, $\omega^2 = 0.043$; expectations/challenge, $F(2, 131) = 4.124$, $p = 0.0183$, $\omega^2 = 0.045$; and motivation/autonomy, $F(2, 131) = 5.351$, $p = 0.0058$, $\omega^2 = 0.061$. Under the Bonferroni reference of 0.0071, prior knowledge and motivation/autonomy remained below the threshold, while expectations/challenge did not. The effect sizes were modest, and no omnibus differences were detected for current-instructor perception, performance/social connection, engagement/curiosity, or course motivation.

6.2. Internal consistency

Table 6 reports the retained alpha estimates without qualitative labels such as “excellent” or “acceptable.” The combined 39-item common subset yielded $\alpha = 0.880$ with 74 complete cases and provides the clearest preliminary consistency result. The HCI and Compilers broad non-course sets yielded $\alpha = 0.938$ and $\alpha = 0.850$, respectively, but both have $k > n$ and combine several theoretically distinct blocks; they are therefore descriptive audits, not evidence of a single latent construct. The course-motivation estimates were 0.785 in HCI, 0.737 in Compilers, and 0.602 in Introduction to Gamification. The last estimate is especially uncertain because it is based on 16 complete cases.

Table 6. Retained Cronbach alpha estimates. The final column limits the evidential use of each coefficient.

Analysis target	Course/data set	Complete n	Items	α	Evidential use
Broad reconstructed non-course set	HCI	16	61	0.938	Descriptive audit only: $k > n$ and multiple theoretical blocks
Broad reconstructed non-course set	Compilers	40	55	0.850	Descriptive audit only: $k > n$ and multiple theoretical blocks
Common cross-course subset	Combined	74	39	0.880	Primary preliminary consistency check; factorial structure still untested
Course-motivation block	HCI	16	15	0.785	Exploratory; small complete-case sample
Course-motivation block	Compilers	40	15	0.737	Exploratory
Course-motivation block	Intro to Gamification	16	15	0.602	Highly uncertain; smallest cohort and complete-case sample

Table 7. Selected non-perfect item-level Pearson correlations with Fisher 95% confidence intervals. No multiplicity adjustment was applied.

Course	n	r	95% CI	Variable 1	Variable 2
HCI	16	0.902	[0.735, 0.966]	Instructor respect	Instructor feedback
HCI	16	0.863	[0.642, 0.952]	Pleasure in taking the course	Taking the course voluntarily
HCI	40	0.916	[0.846, 0.955]	Knowledge of Cronbach's alpha	Knowledge of Gestalt principles
Compilers	62	0.817	[0.713, 0.886]	Knowledge of automata	Knowledge of context-free grammars

6.3. Exploratory associations

Table 7 reports four non-perfect item-level associations. The HCI associations based on $n = 16$ have wide intervals despite large point estimates: instructor respect with feedback, $r = 0.902$, 95% CI [0.735, 0.966], and anticipated pleasure with voluntary enrollment, $r = 0.863$, 95% CI [0.642, 0.952]. The associations involving 40 HCI responses and 62 Compilers responses are more precise, but remain exploratory. Exact correlations of 1.000 found in the small conditional subset were excluded from the interpretive table because they are highly sensitive to sample composition and restricted response patterns.

Table 8 shows the strongest composite-level associations with effective sample sizes and 95% confidence intervals. Expectations/challenge and motivation/autonomy were positively associated in all courses where the pair was examined. The interval for Introduction to Gamification is broad, reflecting its smaller effective sample. These results describe candidate configurations; they do not validate discrete motivational profiles or establish that one construct causes another.

7. Discussion

7.1. Contribution relative to prior work

The study does not claim that semester-start diagnosis, learner modeling, or adaptive instruction is new. Diagnostic-assessment research already frames evidence as part of instructional decision making, and adaptive-learning research already connects learner information to differentiated support [Black and Wiliam 1998, Mislevy et al. 2003, Brusilovsky 2001, Bernacki et al. 2021]. The contribution is the integration of four elements in a computing-education artifact: course-specific perceived prior knowledge, common motivation-aware blocks, preliminary cross-course evidence, and an explicit low-overhead decision protocol. This integration extends prior project work on motivational profiling and learning indicators by adding a baseline that can be revised with later evidence [de Freitas et al. 2024a, de Freitas et al. 2024b].

Table 8. Highest-magnitude exploratory composite correlations by course, with effective sample sizes and Fisher 95% confidence intervals.

Course	<i>n</i>	Composite 1	Composite 2	<i>r</i>	95% CI
HCI	27	Current instructor perception	Motivation and autonomy	0.758	[0.531, 0.884]
HCI	27	Expectations and challenge	Motivation and autonomy	0.735	[0.493, 0.872]
Compilers	88	Expectations and challenge	Motivation and autonomy	0.687	[0.558, 0.784]
Compilers	88	Motivation and autonomy	Course motivation	0.540	[0.373, 0.673]
Intro to Gamification	19	Expectations and challenge	Engagement and curiosity	0.772	[0.489, 0.908]
Intro to Gamification	19	Motivation and autonomy	Engagement and curiosity	0.554	[0.133, 0.806]

The findings also reinforce why diagnosis should be treated as hypothesis generation. Low self-assessed prior knowledge in HCI and Compilers suggests checking for a need for scaffolding, but it does not prove low mastery. High instructor-perception scores provide a favorable relational baseline, but they may also reflect prior experience with the instructor and common-method effects. The selected correlations are semantically plausible, yet their intervals and conditional sample sizes prevent strong claims about stable constructs.

7.2. Heterogeneity and constrained implementation

When substantially different patterns coexist, the response should not be to create a separate syllabus for every student. A feasible design preserves a common core and varies the amount or form of support. For example, a Compilers class may receive a shared visual review of automata and grammars. Students who need more support can use worked examples and a guided practice sequence, while students with stronger prior knowledge can complete an optional parser-design extension. Flexible grouping can be temporary and task-specific, avoiding permanent labels.

The same principle applies to motivation-aware gamification. Students who report high autonomy and curiosity can receive bounded choices of project domain or optional challenge quests, while students who report pressure or low confidence can receive low-stakes rehearsal, clearer success criteria, and private feedback. These options can coexist in a large class because they modify process and support rather than mandatory learning outcomes. The protocol therefore narrows the instructor's decision space; it does not automate intervention assignment.

Institutional constraints remain part of the decision. Under limited instructional time, the priority is one or two actions with broad coverage and low implementation cost. Under a fixed syllabus, personalization should focus on examples, sequencing, feedback frequency, grouping, and optional paths. Under large enrollment, the instructor can use distribution summaries to select a universal response and reserve individual follow-up for students whose diagnostic indicators converge with early performance or participation evidence. A short recheck after two or three weeks is necessary because semester-start responses can change.

7.3. Interpretation of the statistical evidence

The statistical evidence supports cautious diagnostic use, not a claim of full instrument validation. The combined common-item alpha is encouraging, but the broad within-course coefficients mix multiple constructs and, in two cases, use more items than complete cases. Small-sample correlations are accompanied by wide intervals and should be replicated.

Cross-course ANOVA effects are modest and were screened across multiple composites. Consequently, the most defensible conclusion is that the instrument produces interpretable preliminary indicators that can support teacher judgment while its dimensionality, measurement invariance, predictive validity, and intervention utility remain to be tested.

8. Threats to validity

Construct validity. The prior-knowledge block measures self-assessed familiarity rather than demonstrated mastery. Thematic composites were reconstructed for practical diagnosis and have not undergone exploratory or confirmatory factor analysis. Alpha does not establish unidimensionality. The 4-point and 5-point sections were kept separate, but heterogeneous response formats still limit direct cross-block comparisons.

Internal validity. The study is observational and cross-sectional. All student-level indicators were collected by self-report at one time point, so common-method bias may inflate associations. The design does not estimate causal effects of diagnosis, personalization, or gamification on engagement or achievement.

Statistical conclusion validity. Conditional branching, missing responses, and complete-case estimation reduced several effective samples. The broad HCI and Compilers non-course sets have $k > n$, and the smallest correlations and course-motivation coefficient are unstable. The analysis reports confidence intervals for selected correlations, effect sizes for omnibus tests, and a multiplicity sensitivity check. Ordinal estimators, bootstrap intervals, and assumption-robust sensitivity analyses were not included and remain priorities for a preregistered replication.

External validity. The data come from one Brazilian public university, three computing courses, unbalanced cohort sizes, and one academic term. Whether the block structure is invariant across mandatory and elective courses, institutions, study levels, genders, or cultural contexts is unknown. Replication requires larger and more balanced samples.

Pragmatic validity. The proposed actions are theory-informed options, not experimentally verified treatment rules. Their feasibility depends on class size, instructor workload, curricular flexibility, and available support. Future evaluation should document which recommendations instructors select, implementation fidelity, student uptake, and subsequent outcomes.

9. Implications for personalization and gamification

A minimum viable use of the instrument can follow four steps. First, inspect block distributions and combinations rather than rely only on class means. Second, choose one universal action for the most prevalent instructional need. Third, add at most one or two targeted options for contrasting subgroups. Fourth, compare the semester-start hypothesis with short-cycle evidence from formative tasks, participation, attendance, or instructor observation after two or three weeks.

For Compilers, low perceived prior knowledge and high challenge can justify a common review with visual representations, followed by guided exercises for students needing support and an optional extension for students ready for greater complexity. Gamification should emphasize mastery progression, informative feedback, and recovery

opportunities rather than competition. For HCI, low prior knowledge combined with instructional trust can support collaborative interface critiques and low-stakes prototyping before graded evaluation. For Introduction to Gamification, higher autonomy and curiosity can support bounded choices of application domain, optional challenge paths, and reflection on ethical and responsible design.

The same evidence can prevent unsuitable mechanics. High pressure or nervousness argues against public leaderboards or high-stakes competition; low relevance argues for authentic computing cases before adding rewards; high autonomy argues for meaningful choice rather than forced gameful participation. These are conditional design hypotheses that should be revised when behavioral and performance evidence disagrees with the initial questionnaire.

For learning analytics, the instrument can serve as a baseline layer connected to later monitoring data. A teacher-facing dashboard should display distributions, uncertainty, and change over time rather than fixed learner labels. The instructor remains responsible for interpreting evidence, reconciling it with curricular constraints, and deciding whether an adaptation is pedagogically and ethically justified.

10. Conclusion and future work

This paper presented a semester-start, motivation-aware diagnostic instrument used in HCI, Compilers, and Introduction to Gamification, with 245 valid responses. The study situates the instrument within established diagnostic-assessment, adaptive-learning, motivational-design, and differentiated-instruction research and clarifies the boundary between collecting evidence and selecting an intervention.

Regarding RQ1, the questionnaire yielded interpretable preliminary indicators and a combined common-item alpha of 0.880, but the evidence is insufficient to claim full psychometric validation. Regarding RQ2, the course-level distributions showed low self-assessed prior knowledge in HCI and Compilers, high current-instructor perception, and exploratory associations involving expectations, autonomy, pleasure, participation, and conceptual familiarity. Regarding RQ3, the indicators can inform a tiered process that combines a universal core, targeted options, and short-cycle follow-up while preserving common outcomes and accommodating heterogeneous classes.

Future work should harmonize the response scales; add objective prior-knowledge tasks; test factor structure, ordinal reliability, and measurement invariance with larger and balanced samples; compare mandatory and elective contexts; link baseline indicators to longitudinal engagement and performance; and evaluate whether the proposed recommendations are feasible, implemented as intended, and beneficial. The instrument should therefore be understood as an instructor-facing screening and planning aid whose value depends on iterative use with other evidence, not as an automatic profiling or intervention system.

Statement on the use of Artificial Intelligence

OpenAI's ChatGPT was used to support language revision, structural editing, the drafting and refinement of selected passages, literature-oriented suggestions, and preparation of the camera-ready manuscript. All suggested references were independently verified by

the authors. The authors also reviewed and validated all AI-assisted text, analyses, and interpretations and remain fully responsible for the complete content of the article.

References

- Aho, A. V., Lam, M. S., Sethi, R., and Ullman, J. D. (2006). *Compilers: Principles, Techniques, and Tools*. Pearson/Addison-Wesley, Boston, MA, 2 edition.
- Ali, M., Ghosh, S., Rao, P., Dhegaskar, R., Jawort, S., Medler, A., Shi, M., and Dasgupta, S. (2023). Taking stock of concept inventories in computing education: A systematic literature review. In *Proceedings of the 2023 ACM Conference on International Computing Education Research - Volume 1*, pages 397–415. ACM.
- Barbosa, P. L. S., do Carmo, R. A. F., Gomes, J. P. P., and Viana, W. (2024). Adaptive learning in computer science education: A scoping review. *Education and Information Technologies*, 29:9139–9188.
- Bernacki, M. L., Greene, J. A., and Lobczowski, N. G. (2021). A systematic review of research on personalized learning: Personalized by whom, to what, how, and for what purpose(s)? *Educational Psychology Review*, 33:1675–1715.
- Black, P. and Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1):7–74.
- Brusilovsky, P. (2001). Adaptive hypermedia. *User Modeling and User-Adapted Interaction*, 11:87–110.
- Caponetto, I., Earp, J. M., and Ott, M. (2014). Gamification and education: A literature review. In *Proceedings of the 8th European Conference on Games Based Learning*, pages 50–57.
- Chou, Y.-k. (2015). *Actionable Gamification: Beyond Points, Badges, and Leaderboards*. CreateSpace Independent Publishing Platform.
- de Freitas, S. A. A., Farias, M. A. S., Ramos, C. S., Martins, M. V. P., Alves, J. M., and Pinto, L. C. S. (2024a). Crafting personalized learning environments through motivational profiling. In *2024 IEEE Frontiers in Education Conference (FIE)*, pages 1–9. IEEE.
- de Freitas, S. A. A., Ramos, C. S., Bessa, E., Mortari, M. R., and Viana, D. M. (2024b). Implementing neuroscientific principles in gamified software engineering courses. In *2024 IEEE Frontiers in Education Conference (FIE)*, pages 1–9. IEEE.
- Dehghanzadeh, H., Farrokhnia, M., Dehghanzadeh, H., Taghipour, K., and Noroozi, O. (2024). Using gamification to support learning in K–12 education: A systematic literature review. *British Journal of Educational Technology*, 55(1):34–70.
- Deterding, S., Dixon, D., Khaled, R., and Nacke, L. (2011). From game design elements to gamefulness: Defining gamification. In *Proceedings of the 15th International Academic MindTrek Conference*, pages 9–15.
- Dinçer, S. (2020). The effects of materials based on ARCS model on motivation: A meta-analysis. *Elementary Education Online*, 19(2):1016–1042.
- Fang, X., Ng, D. T. K., Leung, J. K. L., and Xu, H. (2024). The applications of the ARCS model in instructional design, theoretical framework, and measurement tool: A

- systematic review of empirical studies. *Interactive Learning Environments*, 32(10):5919–5946.
- Göksu, I. and Bolat, Y. I. (2021). Does the ARCS motivational model affect students' achievement and motivation? a meta-analysis. *Review of Education*, 9(1):27–52.
- Hamari, J., Koivisto, J., and Sarsa, H. (2014). Does gamification work? A literature review of empirical studies on gamification. In *2014 47th Hawaii International Conference on System Sciences*, pages 3025–3034.
- Hunicke, R., LeBlanc, M., and Zubek, R. (2004). MDA: A formal approach to game design and game research. In *Proceedings of the AAAI Workshop on Challenges in Game AI*.
- Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review*, 19(4):509–539.
- Keller, J. M. (2010). *Motivational Design for Learning and Performance: The ARCS Model Approach*. Springer.
- Lavoué, E., Monterrat, B., Desmarais, M., and George, S. (2019). Adaptive gamification for learning environments. *IEEE Transactions on Learning Technologies*, 12(1):16–28.
- Malkewitz, C. P., Schwall, P., Meesters, C., and Hardt, J. (2023). Estimating reliability: A comparison of Cronbach's alpha, McDonald's omega t and the greatest lower bound. *Social Sciences & Humanities Open*, 7(1):100368.
- Mislevy, R. J., Steinberg, L. S., and Almond, R. G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, 1(1):3–62.
- Pellegrino, J. W., Chudowsky, N., and Glaser, R. (2001). *Knowing What Students Know: The Science and Design of Educational Assessment*. National Academies Press, Washington, DC.
- Rodrigues, L. A. L., Palomino, P. T., Toda, A. M., Klock, A. C. T., Pessoa, M. S. P., Pereira, F. D., Oliveira, E. H. T., Oliveira, D. F., Cristea, A. I., Gasparini, I., and Isotani, S. (2024). How personalization affects motivation in gamified review assessments. *International Journal of Artificial Intelligence in Education*, 34:147–184.
- Rogers, Y., Sharp, H., and Preece, J. (2023). *Interaction Design: Beyond Human-Computer Interaction*. John Wiley & Sons, Hoboken, NJ, 6 edition.
- Ryan, R. M. and Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1):68–78.
- Ryan, R. M., Mims, V., and Koestner, R. (1983). Relation of reward contingency and interpersonal context to intrinsic motivation: A review and test using cognitive evaluation theory. *Journal of Personality and Social Psychology*, 45(4):736–750.
- Smale-Jacobse, A. E., Meijer, A., Helms-Lorenz, M., and Maulana, R. (2019). Differentiated instruction in secondary education: A systematic review of research evidence. *Frontiers in Psychology*, 10:2366.
- Tavakol, M. and Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2:53–55.

Tomlinson, C. A. (2017). *How to Differentiate Instruction in Academically Diverse Classrooms*. ASCD, Alexandria, VA, 3 edition.

Werbach, K. and Hunter, D. (2012). *For the Win: How Game Thinking Can Revolutionize Your Business*. Wharton Digital Press.

Zeybek, N. and Saygi, E. (2024). Gamification in education: Why, where, when, and how? A systematic review. *Games and Culture*, 19(2):237–264.