

Inventory of Galilean Transformation of uniform linear motion in position-time graphs

E. B. Merki^{1,*}, S. I. Hofer², A. Vaterlaus¹ and A. Lichtenberger¹

¹Laboratory for Solid State Physics, *ETH Zürich*, 8093 Zürich, Switzerland

²Department of Psychology, *Ludwig-Maximilian-Universität München*, 80802 München, Germany



(Received 28 December 2024; accepted 26 June 2025; published 25 August 2025)

When describing motion in physics, the selection of a frame of reference is crucial. The graph of a moving object can look quite different based on the frame of reference. In recent years, various tests have been developed to assess the interpretation of kinematic graphs, but none of these tests have specifically addressed differences in reference frames. Moreover, existing tests that explore differences in reference frame typically focus on the equivalence principle through written answers, interviews, or simple calculations and vector addition; however, none of these tests evaluate position-time graphs. To address this gap in the research, we developed and evaluated the Inventory of Galilean Transformation of uniform linear motion in position-time graphs (IGT). The IGT consists of 15 multiple-choice items that systematically use position-time graphs of linear uniform motion to assess the understanding of 3 types of Galilean transformations: the identity transformation, the transformation between two (opposing) stationary displaced frames of reference, and the transformation from a stationary into a uniformly moving frame of reference. Herein, we presented the development and validation of the IGT. A total of 532 upper secondary school students in the advanced track participated in the multistage development process. We evaluated the psychometric properties via classical test theory and item response theory. The degree of item discrimination and reliability were within the desired range. The IGT demonstrated good internal consistency ($\omega = 0.88$), and confirmatory factor analysis supported the intended structure of the test. Rasch analysis revealed that the item difficulties were consistent with the increasing complexity of the three different transformations. The IGT also revealed several alternative student conceptions about frames of references, such as confusion between two scenarios of cars crossing versus overtaking, as well as misunderstandings regarding the changing shape of a graph when a uniformly moving object is transformed into a moving frame of reference. In its current form, the IGT serves as a new instrument for assessing students' ability to interpret position-time graphs under the influence of the Galilean transformation, making it suitable for formative or summative assessment in advanced upper secondary education.

DOI: [10.1103/ww1x-kr12](https://doi.org/10.1103/ww1x-kr12)

I. INTRODUCTION

Understanding graphs is crucial in today's data-driven world [1,2], as they are widely used in everyday contexts such as weather reports and stock markets. Mastering graphs requires a solid grasp of the underlying concepts, and physics education plays a vital role in teaching these concepts [3,4]. The ability to read and interpret graphs is also essential for understanding Galilean and special relativity in school contexts, where graphical representations are frequently used to make abstract concepts such as relative motion more accessible. The selection of a frame of

reference (FoR) is central to both theories and becomes visually interpretable in position-time diagrams. Learning about FoRs is often confined to special relativity in advanced upper secondary school tracks [5]. However, kinematics, which is a key subject of every physics course and is typically taught at the beginning of physics curricula in K-12 education, could serve as an excellent opportunity for students to explore how the same motion can be represented differently in position-time graphs, depending on the chosen FoR. The foundational positioning makes kinematics ideal for addressing students' conceptions of FoR [6] early in physics learning.

To assess students' understanding of kinematic graphs, there has been a considerable amount of research. Inventories have been developed that focus on key concepts such as velocity and typical difficulties, such as point/high confusion. However, there is a notable absence of content on changes in FoRs. Among the numerous examples of inventories that test students' understanding are the Test

*Contact author: merkier@phys.ethz.ch

Published by the American Physical Society under the terms of the *Creative Commons Attribution 4.0 International license*. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

of Understanding Graphs in Kinematics (TUG-K) [7], the Kinematics Concept Test (KCT) [8], the Test of Understanding of Vectors [9], the Force Concept Inventory (FCI) [10,11], the Ray Optics in Converging Lenses Concept Inventory (ROC-CI) [12], the Mechanical Waves Conceptual Survey (MWCS) [13], and the Relativity Concept Inventory (RCI) [14,15]. The abundance of these studies may be due to the broad range of skills that students can acquire through exploring graphs. Tests about special relativity focus on important relativistic concepts such as postulates, length contraction, the relativity of simultaneity, and Galilean FoRs [14,16]. These tests include items about FoR [17,18] but rarely include items about uniform movements in an inertial frame [6,19] or graphs. Given that selecting the appropriate reference frame is fundamental to solving any problem in physics, it is surprising that so few studies have examined this topic. To the best of our knowledge, no inventory exists that examines FoRs of linear uniform motion in position-time graphs, although the linear uniform motion position-time graphs are of great importance. A potential reason for this gap in the research might be that, in many cases, the teacher, as the expert, is setting up the FoR for the students, allowing them to focus on solving a specific problem [20,21]. Consequently, students are often overwhelmed when they are required to set up the FoR on their own [17].

To contribute to a better understanding of graphs, an Inventory of Galilean Transformation of linear uniform motion in position-time graphs (IGT) was developed.

This article describes the development and evaluation of the IGT, which systematically assesses the understanding of three increasingly challenging transformations (i.e., T1–T3) based on the Galilean transformation [22]. The first transformation (T1) is the “identity transformation.” Items assigned to T1 assess the necessary basic knowledge about position-time graphs. These items mainly explore the concept of “velocity as a rate” and cover three typical alternative student conceptions described in Beichner’s study with the TUG-K [7], namely, the “graph as picture error,” “slope error,” and “slope/height confusion.” Additional T1 items assess “forward/backward” and “overtaking/crossing” confusion. T2 items assess transformations between stationary FoRs, including translations of the origin along the displacement axis and inversions of the orientation of the displacement axis. Finally, T3 items assess transformations of a position-time graph from a stationary FoR to a moving FoR. Research on student conceptions about FoRs in the context of the Galilean transformation was first performed by Panse *et al.* [6]. By analyzing student answers, the authors identified seven alternative conceptions about FoRs, which will be discussed later. The IGT addresses five of these alternative conceptions.

The following research question is addressed herein: Is the IGT a reliable and valid instrument for assessing

students’ understanding of the Galilean transformation of uniform linear motion in position-time graphs? As part of the validation process, we also analyzed students’ response patterns to explore alternative conceptions reflected in their answer choices.

The IGT was developed and analyzed using feedback from 532 Swiss upper secondary school students. In the following sections, we describe the development and validation of the IGT. The IGT was examined according to the approaches for analyzing multiple-choice questions, which were proposed by Ding and Beichner [23]. We present the psychometric properties and the structural analysis of the students’ responses. Additionally, item response theory (IRT) was applied. The whole validation process was accompanied by guidance from experts in physics education and educational psychology.

The IGT can be used in upper secondary schools as a formative assessment tool to systematically test students’ understanding of the Galilean transformation in position-time graphs. The IGT can also be used to assess prior knowledge about special relativity. For research, we recommend the use of the IGT in combination with a kinematics concept inventory to further evaluate students’ mastery of graphs.

The rest of this study is structured as follows: First, we provide an overview of the background and model (Sec. II); second, we describe our methodology (Sec. III); subsequently, we present the validation of the IGT (Sec. IV) and discuss the findings, student conceptions, and limitations of the study (Sec. V); finally, we present the conclusion (Sec. VI).

II. BACKGROUND AND MODEL OF IGT

Setting up a FoR is fundamental for unambiguously measuring and describing a physical phenomenon. FoRs create a space defined by a coordinate system in which an abstract mathematical description can be applied. In kinematics, students are typically introduced to Cartesian coordinate systems. Learning to establish and transform a coordinate system (i.e., the point of origin and the coordinate axes) is crucial because selecting the optimal system can considerably simplify subsequent mathematical calculations. Furthermore, understanding Galilean transformations of FoRs is important because this information might foster spatial abilities [24,25] and provide necessary prior knowledge for an introduction to special relativity [5].

In their 1994 study, Panse *et al.* [6] identified seven misconceptions regarding FoRs. These misconceptions can be classified into the following categories: (I) the assumption that FoRs can be treated as concrete objects that can catch fire or be destroyed, for example; (II) the localization of FoRs by physical extensions of objects they are “fixed” to, such that, for example, a ball can fall out of such a reference system; (III) the assumption that small bodies located on a larger body can be considered “part of

the frame” of the larger body; this misconception entails the prioritization of relative movement; (IV) the misconception that a phenomenon is associated with a specific frame; this signifies that specific phenomena can only manifest within specific FoRs; furthermore, (V) there is the concept of the real and apparent frame of motion, which pertains to the differentiation between real and apparent movements (e.g., the apparent movement of the train platform observed from the perspective of the moving train is regarded as an illusion); (VI) the concept of describing objects through observation is seen as necessary for perceiving them within a reference system; only what can be observed can be described within that system; and (VII) pseudorelativism, which is defined as the movement of an object being contingent not only on the reference system but also on the manner in which the object is viewed. The IGT addresses misconceptions (I), (II), (V), (VI), and (VII).

It is not possible to address misconceptions (III) and (IV) using graphs. A follow-up study by Ramadas *et al.* [26] found that students often used a “natural” model where velocity becomes an inherent part of the body itself. Understanding and creating FoRs is clearly a major challenge for students. In particular, students have difficulties with selecting appropriate frames [17] and describing the frames for different observers [16,27]. Furthermore, students typically have limited opportunities to practice these tasks, as FoRs are mostly set by teachers in upper secondary school [21].

Many studies have shown that students have difficulty interpreting graphs of linear motion in kinematics [3,7,28]. However, understanding graphs is important, as it is a part of data literacy [2,29] and enhances scientific thinking [30]. The position-time graph in kinematics displays how the position of an object changes over time. An important concept regarding motion in position-time graphs is understanding the “velocity as a rate”; that is, the velocity at a certain time is the slope of the tangent to the position-time graph at that time. In their study related to the TUG-K, Beichner *et al.* [7] identified six common alternative conceptions about uniform and accelerated motion represented in graphs. Relevant to the current study, which is limited to uniform motion, the so-called “graph as picture error” is the situation in which students interpret the graph as a photographic picture of the situation, “slope error” is the situation in which students have difficulty interpreting the slope of a line, and “slope/height confusion” is the situation in which slope values are incorrectly derived from absolute values instead of from gradient triangles. In conclusion, the IGT addresses five misconceptions about FoRs and three misconceptions about graphs that have been previously identified in the literature. Beichner’s TUG-K [7] assessed the concept of the “velocity as a rate” in a predefined FoR; however, it did not assess transformations of the FoR. We therefore defined a set of three transformations, which form the basis of the IGT.

TABLE I. The three transformations, T, each with their own subgroups a and b, are outlined with specific descriptions.

T	Descriptions
T1a	Velocity as a rate—forward vs backward
T1b	Velocity as a rate—crossing vs overtaking
T2a	Transformation between stationary FoRs where the points of origin are separated by $\pm d$, i.e., $x' = x \mp d$
T2b	Transformation between two stationary FoRs where the displacement-axis is reverted, i.e., $x' = -x$
T3	Transformation from a stationary into a FoR moving at velocity $\pm v$, i.e., $x' = x \mp vt$

These transformations are directly derived from the Galilean transformation [22]. The first transformation is the identity transformation; the second transformation is the transformation between two stationary FoRs, including the translation of the origin and the inversion of the position-axis (from x to $-x$); and the third transformation is the transformation from a stationary to a constantly moving FoR, see Table I.

Furthermore, we only considered uniform motions in one direction, which are represented in position-time graphs by straight lines, with the intercept corresponding to the starting point (at $t = 0$) and the slope corresponding to the velocity. We only considered uniform motion in one direction because this phenomenon most clearly illustrates the intended transformation without adding further complexity that we do not want to examine in the test.

We called the transformation T1 identity transformation because the items assigned to T1 assess the understanding of graphs linked to uniform motion within the same FoR. These items assess whether students can distinguish between objects moving forward and backward in a position-time graph. We also defined a subtransformation (T1a), which involves the difference in displaying crossing and overtaking in a position-time graph (see Table I).

The transformations T2 and T3 gradually explore the impact of the Galilean transformation on a position-time graph. T2 examines the transformation between two stationary FoRs. We introduced two subtransformations (T2a and T2b) to assess the displacement and inversion of the axis. More precisely, T2a, as shown in Fig. 1, involves a

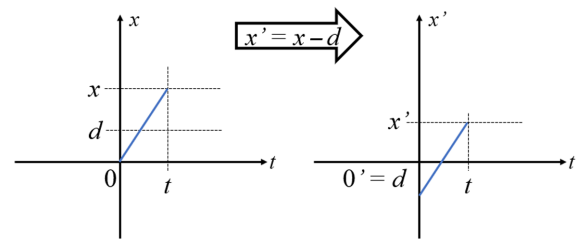


FIG. 1. An example of transformation T2a is the translation of the position axis, which results in a corresponding change in the position-time graph.

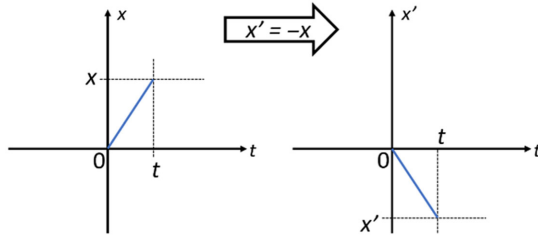


FIG. 2. An example of transformation T2b is the reversion of the displacement axis, which leads to a corresponding change in the position-time graph.

transformation between two stationary coordinate systems $S(x, y, z, t)$ and $S'(x', y', z', t')$ whose points of origin are displaced from each other by the distance $\pm d$ along the x axis, that is, $(x' = x \mp d, y' = y, z' = z, t' = t)$. Geometrically, this transformation is equivalent to a parallel shift of the position-time graph along the x axis.

The transformation T2b (Fig. 2) involves a position-time graph between two coordinate systems $S(x, y, z, t)$ and $S'(x', y', z', t')$, whose axes point in opposite directions: $(x' = -x, y' = y, z' = z, t' = t)$. Geometrically, this is equivalent to converting the slope from v to $-v$, meaning that the graph is reflected over the t axis.

Transformation T3, as shown in Fig. 3, is the Galilean transformation between two coordinate systems $S(x, y, z, t)$ and $S'(x', y', z', t')$, where S' is moving relative to S with velocity $\pm v$ along the x axis ($x' = x \mp vt, y' = y, z' = z, t' = t$), and the point of origin at time $t' = t = 0$ is identical. If a constant motion v_1 in a stationary frame S is transformed into a FoR S' moving with velocity v , the slope of the position-time graph (i.e., v_1) changes (i.e., from v_1 to v_1').

III. METHODS

We used an iterative three-stage process to develop the final multiple-choice version of the IGT. The versions of the first development contained open-ended questions based on the three transformations described above. Student answers and expert feedback were then used to create a closed multiple-choice format of the IGT for the second stage.

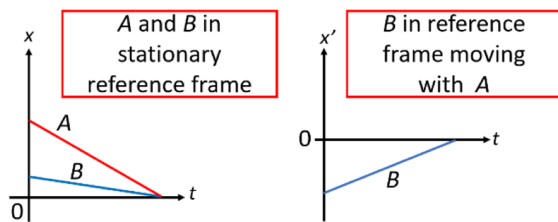


FIG. 3. An example of transformation T3, which involves transforming the graph of object B into the moving reference frame of object A.

Based on test results from the second stage, the questions were reformulated, and the distractors were evaluated to construct the final version of the IGT in the third stage. In the following sections, we describe the data collection, including general information about the participants, the test administration, and the samples and considerations involved in the different stages of development.

A. Data collection

In total, 34 classes involving 11 teachers from 8 different schools participated in the 3 stages of development and validation of the IGT. We collected data from 532 secondary school students in the advanced track who resided in the German-speaking part of Switzerland. In Switzerland, only the top 20% of students are in the advanced track of secondary school. A degree from this school provides access to all Swiss universities. There is no uniform school system in Switzerland, but upper secondary school starts in most areas after the ninth school year. Classes could participate in the IGT validation process if they fulfilled one of the following requirements: the class had recently completed the kinematics course, or the class had completed the kinematics course in the previous year. The average age of the students during the kinematics course was typically between 15 and 16 years.

The test was always presented to an entire class at the same time. The students shared the same classroom and were not allowed to speak with each other. The teacher was only permitted to answer technical questions about the tablets or laptops and to explain terms that the students did not understand. The test versions of the first stage (V1.0–V1.2) were paper-and-pencil-based, and the students who completed them had no specific instruction on FoRs prior to taking the test. The versions V2.0 and V3.0 were completed by students on desktop electronic devices.

Each student had their own device with Internet access, and they could access the test through an online link provided by the teacher. The IGT V3.0 was conducted after a 15-min predict-observe-explain sequence about the Galilean transformation of FoRs to ensure that the students had knowledge about graphs and FoRs.

B. Development process

The iterations and the adaptations of the items are shown in Table II.

1. Open-ended questions V1.0–V1.2

The first three versions (V1.0–V1.2; see Table II) were questionnaires with open-ended questions. The open-ended questions were based on items from the TUG-K and lesson material. The lesson material was provided by a researcher with a Ph.D. in physics education and physics teacher qualifications. An example item is presented in Fig. 4.

TABLE II. The test version, the number of students participating, the types of questions included, and the number of items are listed. The modifications of the tests between two versions of the test are shown in the last column, and the modifications always apply to the next version of the test.

Version	Number of students	Type	Number of Items	Modifications
V1.0	11	Open-ended questions	15	Adding more explanatory tasks
V1.1	6	Open-ended questions	15	Rephrasing of items and adding a drawing part
V1.2	22	Open-ended questions	15	Construction of a multiple-choice questionnaire
V2.0	114	Multiple-choice questions	12	Ceiling effect → some cues in graphs removed and items added for completeness
V3.0	326	Multiple-choice questions	15	

The item shows a picture of a car, with a velocity vector pointing in one direction. Below the car, four position axes are placed, with different points of origin and directions. The students are required to draw the position-time graph for each position axis. The versions V1.0 and V1.1 were completed by 11 students at the end of their first year of physics at a secondary school and by 6 students at the end of the second year. Version V1.2 was completed by 22 students at the beginning of their second-year physics course.

2. Multiple choice test V2.0

Written answers and comments on the open-ended questions were used for the development of the multiple-choice questionnaire. The students’ answers were carefully

and qualitatively analyzed, and the misconceptions reported in the literature were also considered. Consequently, a multiple-choice questionnaire comprising 12 items was constructed (version 2.0). Four items were designed to assess identity transformation (T1), specifically the velocity concept; four items were designed to evaluate the translation of the origin and inversion in the FoR (T2); and four items were designed to analyze the change between two FoRs moving with constant velocity relative to each other (T3). The distractors for T1 are based on well-known student conceptions, which are described in the background section. We further added the “position error,” wherein two cars at the same position are assumed to be overtaking; the “crossing/overtaking error,” which involves confusion between two cars moving in opposite directions and two cars moving in the same direction at different velocities; and the “direction error,” which involves mistakenly believing that a slope smaller than the initial slope indicates that the car is moving backward. The remaining distractors were derived from the responses provided by the students in the open-ended questionnaires, V1.0–V1.2. The distractors from the open-ended questions were classified according to the following categories. The term “translation error” is used to describe instances wherein the graph is translated in the incorrect direction. A “mirroring error” describes cases wherein the graph is either point mirrored at the origin, line mirrored with respect to the position axis, or transformed using another axis or line. “Shift error” occurs when a graph is transformed into the correct coordinate system, but the slope of the movement in the new coordinate system is not accurate. The “translation error” and “mirroring error” align with the misconception (VI) proposed by Panse *et al.*, which posits that students ascertain motion on the basis of visual appearance rather than by transforming velocities. The “shift error” is analogous to the alternative conception (I) proposed by Panse *et al.* [6], which occurs when students perceive the motion to be attached to the frame. The frame is altered, but the slope of the motion remains unadjusted. The “origin error” is the belief that a graph

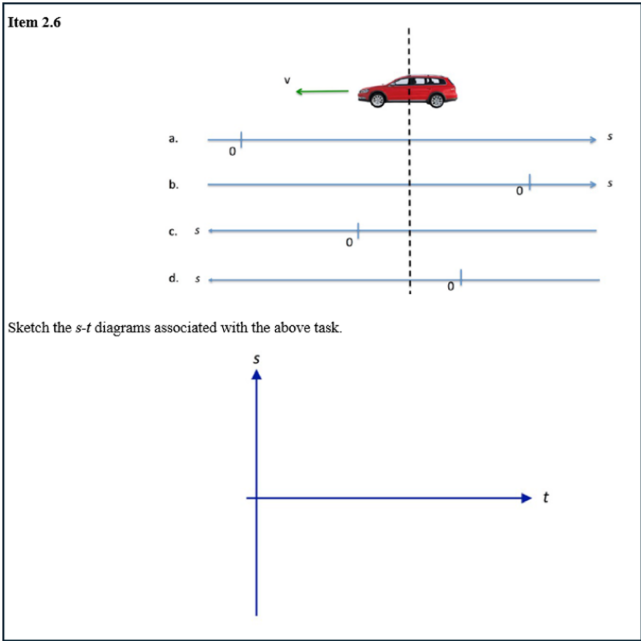


FIG. 4. An example of an open-ended question from versions V1.0–V1.2 of the IGT, which was designed to explore student conceptions.

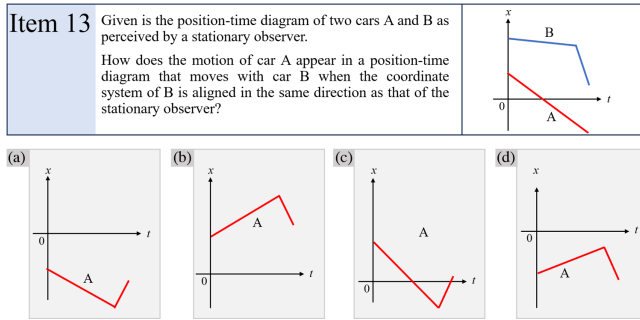


FIG. 5. Item 13 from IGT V3.0 with the correct answer (a) presents three potential distractors: (b) “transformation error,” (c) “origin error,” and (d) “shift error.”

intersecting the time or displacement axis must do so in all possible coordinate systems. This error is illustrated in Fig. 5, distractor C. The “orientation error” is the transformation of a graph without reversing the axes or vice versa.

Additionally, one distractor has been identified that partially aligns with a combination of errors and is exclusive to (T3) items. A “transformation error” occurs when the graph is transformed into the coordinate system of the incorrect vehicle. For example, the graph of car A is transformed into a coordinate system traveling with car B and vice versa. The transformation error is analogous to the alternative concept (II) described by Panse *et al.* [6], wherein students fail to recognize that a FoR is independent of the body.

The alternative conception (V), as described by Panse *et al.* [6], did not manifest any specific misconceptions. However, the IGT items required the students to transform the graphs from their favored FoR to another, thereby challenging the alternative conception of real and apparent motion. The alternative conception of pseudorelativism (IV) described by Panse *et al.* [6]—i.e., the conception that a physical description of a phenomenon is not unambiguous but depends on how it is viewed—did not emerge in responses to the open-ended questions. Rather, this alternative conception is a prerequisite for even posing questions. In general, all of the IGT items present a motion accompanied by a graph and request the representation of the same motion with an alternative graph. The test was validated by a researcher and lecturer with a Ph.D. in physics education and physics teacher qualifications, a Ph.D. student in physics education with a physics teacher qualification, and an educational psychologist. In this test version, each item was designed to test exactly one transformation. The test version V2.0 was completed by 114 students (82 female, 32 male) across four classes and two teachers from two different schools. The students were in their first-year physics course. The internal consistency was 0.85. A closer analysis revealed that the item difficulty of the T2 items was similar to that of the T1 items and

that the T3 items were easier to solve than the T2 items. Of the 12 Rasch parameters, 7 ranged from -0.5 to 0.5 , with an information curve with a negative maximum between -2 and 1 .

3. Final version of the multiple-choice test IGT V3.0

Because the T2 and T3 items were less challenging than intended, we decided to adjust multiple items in version V3.0. A new response was added for item 2 to test whether students could distinguish between a car crossing and a car passing. A content analysis of a formative assessment using peer instruction revealed that not all students understood the difference between the maneuvers [31]. A new item, item 4, was added to retest overtaking versus crossing. Item 4 of V2.0 became item 5 of V3.0. This process consecutively changed the number of items. The next four items testing T2 were modified by reformulating the question. In addition to the reformulation of the question for the second, third, and fourth items, T2 items with translation and inversion transformations were added to make them more difficult. In version V2.0, four items tested transformation T3. We decided to include two more items for completeness, where both cars move backward. Furthermore, we removed some cues from the items. This resulted in a total of 15 items in version V3.0 of the IGT. Table III shows all the items and the assigned transformations. Some items (8, 9, 14, 15) test two transformations (translation and inversion). The idea was that two transformations stacked on top of each other would increase the difficulty. To add redundancy, each transformation was applied once for forward motion and once for backward motion. This version of the IGT includes five items assessing the identity transformation (T1). Of these five items, two assess whether students understand the difference between overtaking and crossing when the cars move backward (T1a). A graph is shown with four multiple-choice answers formulated as short sentences. Items 6–15, which assess transformations T2 and T3, have the following structure: a graph with a question is given, and students select one out of four graphs as their answer. The distractors for these items were constructed using the answers drawn by the students in the open-ended questions of V1.0–1.2. Items 14 and 15 include distractors based on all the errors mentioned above. Before the test was administered to 326 high school students, it was validated by a physics professor, a lecturer in physics education with a physics teacher qualification, a researcher and lecturer with a Ph.D. in physics education and physics teacher qualification, two Ph.D. students in physics education research with a physics teacher qualification, and an educational psychologist.

The psychometric properties, structural analysis, and Rasch scaling of the test version V3.0 are presented in detail in the evaluation section. The IGT was developed in

TABLE III. Detailed test structure of the IGT V3.0: the underlying transformation and the distractors/student conceptions associated with each item. The letters in parentheses (A, B, C, and D) refer to the answer choices in the test and correspond to the response frequencies presented in Table V.

Item	Transformation	Distractors/student conceptions		
		1	2	3
1	T1a	Slope error: negative position means moving backward (A)	Position error: end point at the same position does not mean crossing (C)	Position error: same end point means same displacement (D)
2	T1b	Crossing/overtaking error: overtaking is not crossing (A)	Slope error: negative position means moving backward (B)	“None of the statements are correct” (D)
3	T1	Crossing/overtaking error: overtaking is not crossing (A)	Slope/Height confusion: unable to find directions from graph (B)	Slope error: negative position means moving backward (D)
4	T1a	Crossing/overtaking error: overtaking is not crossing (A)	Direction error: change between slope of same sign means change of direction (B)	Slope error: positive position means moving forward (C)
5	T1a	Direction error: change between slopes of the same sign means a change of direction (A)	Crossing/overtaking error: overtaking is not crossing (B)	Slope error: negative position means moving backward (D)
6	T2a	Translation error: translation into the wrong direction (B)	Mirroring error: point mirroring at the origin (C)	Mirroring error: mirroring on position axis (D)
7	T2b	Mirroring error: mirroring across a line parallel to the position axis (B)	Mirroring error: mirroring across some axis parallel to the position axis (C)	Orientation error: wrong translation, no mirroring (D)
8	T2a/T2b	Orientation error: wrong translation, no mirroring (B)	Orientation error: correct translation and point mirroring (C)	Orientation error: correct translation and mirroring on the position axis. (D)
9	T2a/T2b	Orientation error: correct translation, mirroring on a line parallel to the time axis (A)	Orientation error: correct translation, no mirroring (B)	Orientation error: translation in the wrong direction, no mirroring (D)
10	T3 forward	Origin or graph as picture error: incorrect slope, too steep (B)	Shift or graph as picture error: incorrect slope (C)	Shift error: correct sign of slope, but too steep (D)
11	T3 backward	Shift error: correct sign of slope, but too steep (B)	Origin error: same starting point as original graph (C)	Transformation or graph as picture error: correct answer mirrored at time axis (D)
12	T3	Transformation error: shifted onto the wrong object (A)	Graph as picture error combined with origin error: same starting point as original graph and incorrect slope (C)	Shift error: correct transformation incorrect slope. (D)
13	T3 backward	Transformation: shifted onto the wrong object (B)	Origin error: passing through the origin (C)	Shift error: correct translation, incorrect slopes (D)
14	T3	Transformation error: no change of orientation (A)	Shift error: incorrect slope (D)	Orientation error: incorrect slope and no change of orientation (B)
15	T3	Transformation error: no change of orientation (B)	Orientation error: incorrect slope and no change of orientation (C)	Shift error: incorrect slope (D)

German. The nonvalidated English translation of the test is provided in Supplemental Material [32].

C. Student sample

The final 15-item version V3.0 was completed by 326 students (183 females and 143 males) across 11 teachers at 8 different schools (for a total of 27 classes), including 48 students (27 females and 19 males) in their second year of the physics course and 280 students (156 females and 124

males) in their first year of the physics course. The test data used herein originated from students who fully completed the test and consented to the use of their data. A majority of the 280 students (86%) completed the test directly after a short introduction to FoRs and a brief 15-min predict-observe-explain sequence on the Galilean transformation of FoR in the kinematics section. The other 46 students (14%) completed the IGT after an introduction to FoRs. We included only students who reported that their data were

trustworthy. The response options in the multiple-choice test were displayed in random order. Therefore, we could not know if a student consistently chose the first answer. Nevertheless, the data do not indicate any such pattern since randomly choosing one answer with a probability of one-fourth of 15 questions would result in a guess score of 3.75. If we assume that uninterested students responded quickly by randomly selecting answers, we would have expected an accumulation of short answer times with three to four correct answers, which we did not observe.

D. Data analysis

We used two approaches suggested by Ding and Beichner [23] for the data analysis of multiple-choice questions: classical test theory, including the evaluation of psychometric properties and confirmatory factor analysis, and IRT. The programming language R [33] was used to conduct the data analysis.

Classical test theory examines whether a test is reliable and discriminating. Item analysis includes three measures: item difficulty, the discrimination index, and the point-biserial coefficient. The item difficulty is the ratio of correct answers to the total number of answers. The discrimination index shows how well an item discriminates between low- and high-achieving students [34]. The point-biserial coefficient is defined as the correlation between the item scores and the total scores [35].

Confirmatory factor analysis (CFA) was used to test whether the transformations (T1, T2, and T3) mapped onto the items of the IGT—that is, we tested whether the theoretical model based on our defined transformations can be found in our data. As we constructed the IGT items on the basis of different transformations, the patterns of correlations between the items (indicator variables) should reflect the different transformations as latent factors. We first specified the CFA model for the IGT; then, we defined the structural equations. Finally, we chose the iterative method. The IGT is a multiple-choice questionnaire with dichotomous responses on a continuous scale. One of the four response options can be selected for each item. A correct answer is worth one point, and an incorrect answer is worth zero points. Therefore, the weighted least squares method with robust standard errors and a mean- and variance-adjusted test statistic (using a scale-shifted approach; *wlsmv*) was chosen. The IGT has a minimum of at least four items per factor and a sample size larger than 200; therefore, a reliable model estimate can be made [36]. The model syntax from the *LAVAN* package [37] in R was used to conduct the confirmatory factor analysis. The goodness of fit of the model of the CFA was examined via the comparative fit index (CFI), the Tucker-Lewis index (TLI), and the root mean square error of approximation (RMSEA). The CFI assesses the fit of the model by comparing the sample covariance matrix to the hypothetical model while correcting for the sample size issues

associated with the chi-square test. CFI values range from zero to one, with higher values indicating a better fit [38]. The TLI, also known as the non-normed fit index, evaluates the improvement in model fit by comparing the chi-square value of the hypothesized model to that of the null model. According to the literature, TLI values larger than 0.90 [39] or 0.95 [40] indicate a good fit. The RMSEA is a fit index of the standard deviation of the residuals between observed and modeled correlation matrices [41]. We used the *semPlot* package of R to construct the path diagram [42].

The second approach we used for the analysis was IRT. Students' latent abilities and item characteristic parameters can be estimated via IRT [43]. The items of the IGT are based on the concept of velocity as a rate [7] and consist of graphs; all the items are measured on a continuous scale. We used the one-parameter Rasch model for dichotomous items, which was developed in 1960 [44]. It is the simplest model of the Rasch model family [45] and represents a specific method within IRT. Unidimensionality [46] was tested for the dichotomous Rasch model through principal component analysis (PCA) of the residuals. For our model, 37% of the variance in the student responses could be explained by estimates of students and item locations on the logit scale. The PCA of the residuals revealed the largest eigenvalue close to the critical value of 2.00 [47]. The goal of this analysis was to show that the item difficulties were distributed according to their specification into easy items for concept T1, medium to difficult items for the transformation T2, and difficult items for the transformation T3, as in Ref. [48]. The item difficulty parameter, which is obtained via the Rasch model, does not have the same meaning as in classical test theory. The item difficulty is the ability value of a 50% probability of a correct response. Therefore, if an item is more difficult, a higher ability level is required for a student to obtain a 50% probability of the correct answer. We used the *TAM* library in R to perform Rasch analysis [49], and we used the *eRm* package to construct the item characteristic curves (ICCs) and the test information curves (ITFs) [50]. We further used the *PSYCH* package [51] for statistical analysis and the *GGPLOT2* package [52] for plotting.

IV. VALIDATION OF THE IGT

In what follows, we present our analysis of the inventory's use at the upper secondary school level, including psychometric properties and a discussion of the student answers, results from the factor analysis, and findings from the IRT analysis.

A. Psychometric properties and student answers

Table IV shows the mean psychometric properties and the desired ranges [23]. For individual items, the item characteristics and desired ranges are presented in Table V.

TABLE IV. The mean psychometric properties of the IGT.

Statistic	IGT value	Desired range
Mean item difficulty	0.55	[0.3, 0.9]
Mean discrimination index	0.42	≥ 0.3
Mean point biserial coefficient	0.43	≥ 0.2

The average test score was 8.2, which is close to the optimum value of discrimination halfway between the random guessing score of 3.75 and the maximum possible score of 15.

The distribution of item difficulty was as follows: 20% of the items fell within the difficult range (0.00–0.35), 33% fell within the moderately difficult range (0.35–0.60), 40% fell within the moderately easy range (0.60–0.85), and 7% fell within the easy range (0.85–1.00).

This distribution is consistent with Doran’s [53] suggestions of having 15% of the items in each of the two extreme categories and 35% of the items in the two moderate categories. However, the easy items are slightly under-represented, with a higher proportion of difficult items instead. To estimate the internal consistency of the test, we applied Kuder-Richardson’s Formula 20 and obtained a value of 0.69, which indicates relatively low reliability. However, this value has limited significance because the test does not have a strictly unidimensional structure (see the “Factor analysis” section). For the five items assessing identity transformation (T1), the internal consistency is 0.89; for the four items testing the transformation T2 between two stationary FoRs, the internal consistency is

0.69; and for the six items testing the transformation T3, the internal consistency is 0.83. The low internal consistency of the items assigned to T2 might be attributed to the smaller number of items. The point-biserial correlation and discrimination indices were calculated to assess the quality of each item. The point-biserial coefficient must be greater than 0.2 [23] to be considered reliable. Except for item 15, all the items exceeded the threshold. The discrimination index assesses the ability of a test item to discriminate between students with low and high scores by evaluating the difference between the top quartile and the bottom quartile of respondents; a value greater than 0.3 indicates good reliability [23]. Only items 14 and 15 were below this threshold. Ferguson’s delta was $\delta = 0.96$, which provided an additional measure of the discriminatory power of the entire test and confirmed that the IGT can effectively discriminate among students at different levels. Considering that item 15 did not meet the desired criteria for the point-biserial coefficient and the discrimination index, we assume that this item was too challenging for students to select the correct answer. One distractor was nearly correct, differing only by a minor displacement between the graph and the axis; therefore, the difference between the two options may have been too difficult to discern on a tablet. With respect to item 14, the low item difficulty due to the challenging inclusion of an axis inversion (T2b) at T3 might have led to a reduced discrimination index. If only 15% of the students can solve an item correctly, the maximum value for the discrimination is lowered to 0.60, and even a small number of false positives or negatives (e.g., due to guessing) can

TABLE V. Psychometric properties and response frequencies for the IGT. The table presents item-wise values for difficulty, point-biserial coefficient, and discrimination index. Asterisks indicate values below thresholds, which are discussed in the main text. The last column shows the number of students who selected each answer option (A, B, C, D); the correct answer is highlighted in bold.

Items	Difficulty	Point-biserial coefficient	Item discrimination index	Choice ($N = 326$)			
				A	B	C	D
Desired value	0.3–0.9	≥ 0.20	≥ 0.30				
1	0.84	0.48	0.37	24	273	5	24
2	0.66	0.47	0.50	54	17	215	40
3	0.86	0.49	0.32	17	6	281	22
4	0.80	0.56	0.47	36	4	24	262
5	0.75	0.50	0.49	0	63	243	20
6	0.35	0.31	0.32	115	175	27	9
7	0.75	0.48	0.46	245	13	31	37
8	0.73	0.46	0.48	238	37	21	30
9	0.56	0.45	0.51	89	28	182	27
10	0.39	0.50	0.59	127	96	94	9
11	0.27*	0.40	0.39	89	30	52	155
12	0.43	0.53	0.63	19	141	144	22
13	0.32	0.43	0.47	104	68	93	61
14	0.15*	0.30	0.23*	79	93	48	106
15	0.38	0.17*	0.10*	124	110	56	36

significantly reduce the value. A detailed analysis of the items and selected distractors is presented in the next paragraph.

In addition to the item characteristics, Table V also shows the distribution of the students' answers across the various response options. For items 1–5, which assess the transformation T1, most of the students selected the correct answers. However, it is worth taking a closer look at the most frequently selected distractors. For item 2, 17% of students chose answer A, which states that the cars are overtaking instead of crossing. For item 4, in which the cars are overtaking, only 11% of students selected the crossing distractor. For item 5, in which the cars are also overtaking, but the slope of one of the graphs is almost zero, 19% of students chose the wrong answer attributed to crossing. The T2 section of the IGT, which assesses translations and axis inversions of stationary coordinate systems, includes items 6–9. Items 7–9 were solved correctly by most students. However, for item 6, which tests a translation of the coordinate system (T2a), a remarkable majority of the students chose the translation in the wrong axis direction, whereas only 35% selected the correct translation along the x -axis. The significantly lower percentage of correct answers compared with items 7–9 prompted a rethinking of the wording of item 6, which reads, “The position-time diagram of cars A and B is given. Both cars have the same start and end point. Which position-time diagram describes the motion if the point of origin of the displacement is now at the end point of the same displacement?” We suggest the following rephrasing: “Given is the position-time graph of two cars A and B. Both cars have the same start and end point. Which position-time graph describes the motion if the end point of the path becomes the origin in the transformed coordinate system?” Owing to the ambiguous wording in item 6, one of the distractors could be interpreted as a correct answer, leading to false negatives.

For item 7, which involves an axis inversion (T2b), answer C (which reflects an incorrect mirroring of the axis) and answer D (which presents a translation instead) were the most frequently selected distractors, chosen by 10% and 11% of students, respectively. Items 8 and 9 combined translations (T2a) and axis inversions (T2b) in a similar way. However, item 8 was solved correctly much more often than item 9. Regarding the distractors in item 8, 6% and 34% of students chose answers C and D, respectively, which contained correct translations of the graph but incorrect graph reflections. Only a few students chose distractor B, which included an incorrect translation and no reflection at all. With respect to the distractors in item 9, most of the students who answered incorrectly selected option A, in which the graphs were correctly reflected but incorrectly translated. The absence of this kind of distractor in item 8 (correct reflection, incorrect translation) might have possibly led to the higher rate of correct responses. Items 10–15 of the IGT assess T3, including

transformations of FoRs moving at constant velocity. These items were generally more difficult than items assessing T1 and T2. Only items 10, 13, and 15 were solved correctly by a majority of the students. The most frequently selected distractors for item 10 were answer B, which shows the graph of car B in the stationary system relabeled as car A, and answer C, which displays an incorrect displacement between cars A and B, where the cars move away from each other instead of approaching each other. Item 11 asks students to transform the graph of car B into the moving FoR of car A. Nearly half of the students selected the incorrect answer D, in which the graph of car A in the stationary system is relabeled as car B. Another 16% chose distractor C, in which car B's graph has the correct slope but the same starting point as car A in the stationary system. Finally, 9% chose distractor B, in which the displacement between the cars is incorrect. Even though items 10 and 11 covered the same transformation, item 11 was slightly more difficult. One possible reason might be the direction of the cars: in item 10, the cars move forward, whereas in item 11, they move backward, which is typically a more challenging concept for students to grasp. An additional challenging aspect of item 11 could have been the switch in transformations between cars A and B: item 11 required transforming the graph of car B into the constantly moving FoR of car A, whereas in item 10, the graph of car A had to be transformed to the constantly moving FoR of car B. Item 12 is similar to items 10 and 11, except that car B changes from a slower to a higher constant velocity, resulting in a kink in the graph. The graph of car B had to be transformed into the constantly moving FoR of car A. Similar to items 10 and 11, the most frequently selected distractor was answer C, which simply shows the graph of car B relabeled as car A, but without any transformation applied. Distractors A and D were selected by only a small number of students; these responses show correct slopes but an incorrect reflection over the t axis (A) and an incorrect displacement (D). Item 13 also features a kink in the graph of car B. In contrast to item 12, the graph of car A, which moves at constant velocity, has to be transformed in the FoR of car B. All distractors were frequently selected. The most common incorrect choice was distractor C, which is closest to the original graph of car A. This distractor shows a qualitatively correct progression of the slope but an incorrect displacement. Distractors B and D, both showing graphs with incorrect reflections over the t axis, were chosen by about 20% of students each. The graph in distractor D also has incorrect displacements. Items 14 and 15 were considered the most challenging, as they also include axis inversions (similar to transformation T2b). All distractors in item 14 were selected more frequently than the correct answer. The most common incorrect choice was answer D (33%), in which the direction of motion for car A is correct (correctly considering the axis inversion), but the

displacement with respect to the time axis in the moving FoR is incorrect. Distractor A, which shows the correct displacement of the graph but neglects the axis inversion, was selected by 24%. Finally, distractor B was selected 29%, where both the displacement and the reflection of the graph over the t axis are incorrect. For item 15, distractor B was selected by 34% of students, where the displacement of the graph for car B is correct, but the axis inversion is neglected. Distractors C (17%) and D (11%) both show graphs with incorrect displacements, and in the case of answer C, there is also an incorrect direction of motion (inversion neglected). For item 15, a considerably greater number of students chose the correct answer than did for item 14. There is no obvious explanation for this difference, as the items are similar in design. The similarity of distractor D to the correct solution in item 14 may have influenced this outcome. If we count all the answers of the students that include a correct displacement, regardless of axis inversion, the totals are 39% for item 14 and 72% for item 15. Clearly, the additional axis inversion increased the difficulty compared with items 10–13. This additional challenge, which might have been overlooked by some students, might explain the relatively low point-biserial coefficients and discrimination indices of items 14 and 15.

B. Factor analysis

We performed CFA to assess whether the intended structure of the IGT was represented in the students' answers. The model had three factors that aligned with our transformations. The first five items were tested on the same transformation T1; therefore, all the items were grouped into one factor, as were the next four items testing

transformation T2 and the last six items testing transformation T3. The scaled chi-square test of model fit yielded a chi-square value of 106.989, d.o.f. (87), and $p = 0.072$. The RMSEA scaled value was 0.027, with a 90% confidence interval of [0.000, 0.042], indicating a good model fit (values below 0.05 are considered good). The probability of the RMSEA being less than 0.05 was 0.996, further supporting the good fit of the model. The CFI for this model was 0.982, which exceeds the commonly accepted threshold of 0.95, suggesting an excellent fit. The TLI for this model was 0.978, which exceeds the commonly accepted threshold and thus also suggests a good model fit.

Figure 6 displays the latent variables and the item loadings on the factors. All loadings were considerably high (> 0.5) except for items 6, 14, and 15. Possible issues with these items have already been discussed. We suggest slight revisions to these items for future applications of the test (see also the “Limitations” section). The factors exhibit substantial correlations, ranging from 0.35 to 0.54.

C. Item response theory

We conducted a Rasch analysis to further assess the quality and reliability of the IGT. Additionally, we aimed to verify whether the item difficulties aligned with their intended specification: easier items for transformation T1–“identity transformation,” medium to difficult items for transformation T2–“translation and rotation of FoR,” and difficult items for transformation T3–“transformation from a stationary into a moving FoR.”

The R function using the Green and Yang method [54] for dichotomous data calculates the reliability via the tetrachoric matrix ($\alpha = 0.81$, $\omega = 0.87$). The

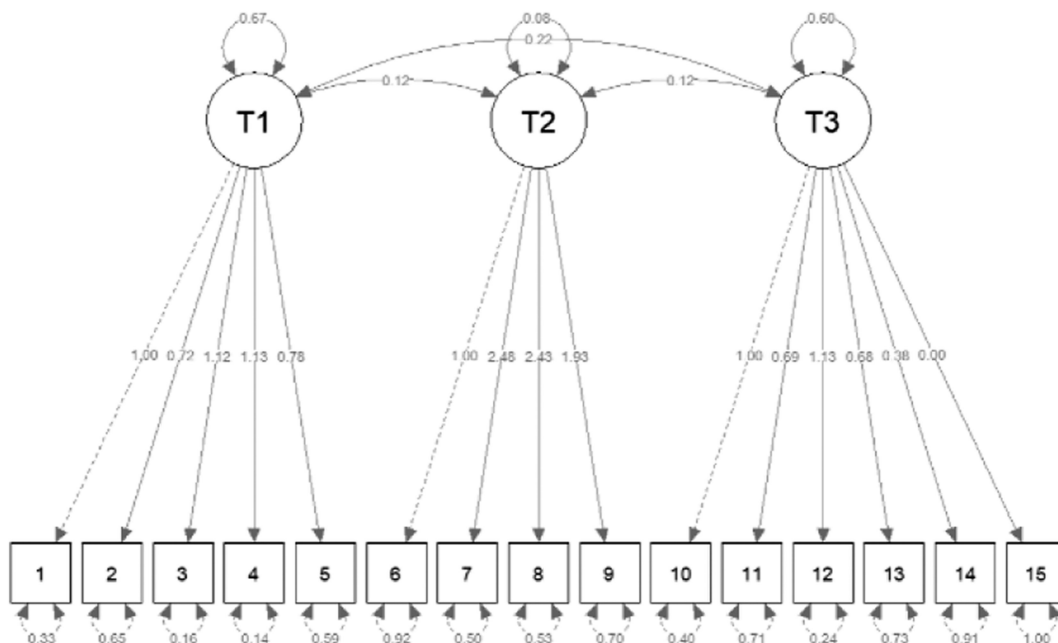


FIG. 6. CFA for IGT with 15 items; the numbers correspond to the item number.

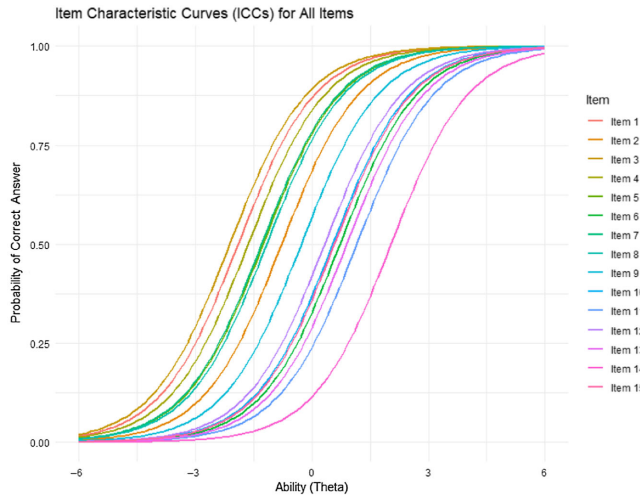


FIG. 7. The item characteristic curves for all items of the IGT.

IRT 1PL (one-parameter logistic) model was used, and it converged after 13 iterations. We calculated item difficulty and plotted the probability for each item against ability. The ICCs of these items are shown in Fig. 7.

The ITFs for the items have maxima ranging from -2.5 to 2.0 (see Fig. 8), indicating the information that each item provides about a person parameter.

Differential item functioning was not different between the two groups or between the two levels of experience. Warm's mean weighted likelihood estimate (WLE) reliability was 0.67 [55]. The expected *a posteriori* measure (EAP) reliability was 0.68 . The EAP estimates the most likely ability level for each participant on the basis of their observed pattern or the level at which their response string best fits the Rasch model. The items with the lowest scores were items 1–5, which were assessed at T1. These items had a mean parameter value of -1.54 . Items 6–9, which assess transformation T2, were of medium difficulty, with

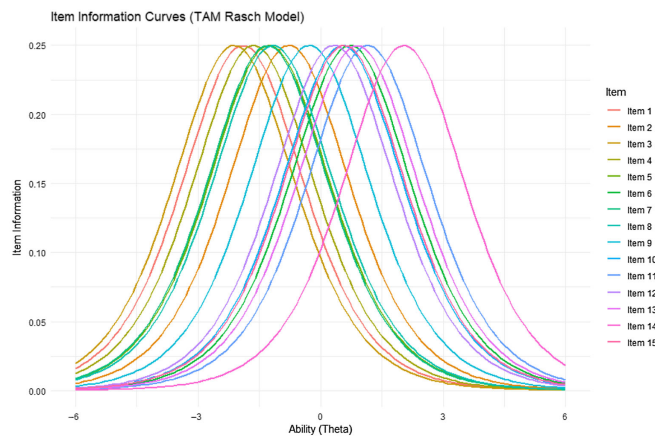


FIG. 8. Test information curves for the 15 items of the IGT are displayed. The maximum item information ranges from -2.5 to 2 .

a mean parameter value of -0.51 . Items 10–16, which assess transformation T3, had the highest mean difficulty at 0.92 . A high infit value indicates that an item is unpredictable, whereas a low infit value suggests that the item does not contribute meaningful information to the survey. A high outfit value indicates the presence of many outliers. The infit values ranged between 0.88 and 1.23 , whereas the outfit values ranged between 0.74 and 1.37 .

Item 15 stands out with a high infit value of 1.23 and a high outfit value of 1.37 . The Rasch analysis confirmed that the item difficulties were distributed as expected, from easy to difficult, across transformations T1 to T3. Except for item 15, all infit and outfit values fell within the acceptable range.

V. DISCUSSION

The aim of the present study was to develop and validate the Inventory of Galilean Transformation of uniform linear motion in position-time graphs (IGT). Furthermore, we explored alternative student conceptions through its administration. The IGT is a multiple-choice test that systematically assesses the transformation of graphs representing uniform motion between frames of reference (FoRs). The IGT is designed for advanced upper secondary and high school tracks, and the presented graphs are less complex than those in other kinematics tests [7]. Moreover, the test avoids the use of symbolic-mathematical representations for advanced understanding. Prior knowledge for students before completing the test should include a basic introduction to graphs and the display of uniform motion in a graph. Although there are many well-known difficulties associated with the Galilean transformation [6], the IGT is limited in its ability to imply position-time graphs of uniform motion. The IGT was presented to a sample of 326 advanced upper secondary school students who had completed a kinematics course.

The results showed high internal consistency and acceptable psychometric properties in the desired range. CFA revealed that the three transformations covered in the IGT had a good model fit.

Rasch scaling indicated good differentiation between the difficulties of the different transformations. Additionally, the ITFs indicated that the differentiation among highly competent students was less effective. One potential explanation for this finding is that the items assessing transformations from a stationary into a moving FoR were too challenging for upper secondary school students. The infit values indicated that there is a balance between items that cover the different correlated scales [56].

The one-dimensional Rasch model demonstrated satisfactory item fit values, although the CFA indicated that the multifactorial model yielded a superior fit, suggesting that the test can be utilized to derive both an overall score

(supported by Rasch) and multiple scores for the individual dimensions (supported by the CFA).

The distractors selected by the students provide insight into the students' conceptions and learning process. The items assessing the first transformation, which involve basic knowledge about position-time graphs for uniform motion, revealed that some students have difficulty distinguishing between situations wherein a car is overtaking another car, wherein a car is moving toward another car, and wherein both cars are crossing each other. The items assessing the second transformation involved the translation between two stationary FoRs and/or the inversion of their axes. For item 6, almost half of the students selected the distractor wherein the graph was translated in the wrong direction. Further evaluation of the item revealed that the question could be interpreted in a way that made the distractor with the wrong translation appear correct. For the IGT in Supplemental Material [32], we corrected this ambiguity. All other items related to this transformation were solved correctly by a majority of the students. For the third transformation, students had to select a graph for car A in a FoR that moves with car B. The questions included both graphs in a stationary environment. We observed an alternative conception similar to the graph-as-picture error [7]. The item can be described as follows: two graphs of constantly moving objects A and B are given in a resting FoR. The goal is to transform the graph of B in the moving FoR of A. The students chose the distractor wherein the graph of B had the same shape as in the question. We believe that this is because the graph was the only one in the solution that resembled the graphs shown in the question. The other distractors that were mostly chosen by the students for the other items and transformations included the following errors: translation error, mirroring error, slope error, transformation error, and orientation error. In summary, the student difficulties we found regarding the third transformation (items 10–15) correspond well to the misconception found in relative motion research where students have difficulties visualizing trajectories in different FoRs [57].

The average psychometric properties of the IGT exceed thresholds; nevertheless, a detailed review of the items shows that item 15 has five out of seven indices below the acceptable thresholds. Furthermore, we also obtained feedback from the teachers who participated in their classes that the lines in the position-time graph of the four different answers were difficult to distinguish. Therefore, we suggest not using that item anymore. We provide the psychometric properties of the IGT with 14 items in the Appendix.

A. Outlook

One purpose of this instrument was to highlight the different levels of understanding for a systematic approach

to learning about changes in FoR for position-time graphs under the Galilean transformation. The reported results are important steps toward a robust instrument, but the IGT will certainly be improved by modifying some items or applying it to additional samples. However, given its sufficient psychometric properties, it can be used as an instrument to measure learning gain resulting from instruction. As we found no existing test that systematically assesses knowledge about Galilean transformations in position-time graphs, we suggest evaluating instruction and practicing this content. It might be useful to teach students about the transformation of FoRs after the introduction of position-time graphs in kinematics at the beginning of the upper secondary school course. Knowledge about FoRs and transformation is an important prerequisite for special relativity, which is typically taught later in the physics curriculum. Compared to existing inventories that focus on Newtonian mechanics or special relativity, the IGT could contribute to systematically addressing conceptual understanding of Galilean transformations in position-time graphs—a topic that is often underrepresented in current assessments. We suggest combining the IGT with the TUG-K [6] to include the concept of the transformation of FoRs into kinematic graphs; another interesting possibility is to use items from the IGT as a prior knowledge component in the RCI [13]. It might also test transfer knowledge at the university level, where students transfer the knowledge from special relativity back to graphs in kinematics. In this context, however, the validity of the IGT still needs to be established. We have two ideas for future research. First, it could focus on the application of the Galilean transformation to representations other than graphs, such as Minkowski diagrams or vector sketches used to demonstrate simultaneity, time dilation, or length contraction. New items can be created and added to the IGT. Second, future research could focus on rotating FoRs (e.g., the Earth) or accelerated FoRs and the implications for graphs or vectors. Furthermore, it would be interesting to explore students' ideas about forces and the paths of objects as observed inside and outside rotating FoRs.

B. Limitations

There are several points to consider, despite the promising possibilities offered by the development of the IGT. We suggest applying the test without the last item and

TABLE VI. The mean values for item difficulty, the discrimination index, and the point biserial coefficient, along with the thresholds from [23] for the IGT.

Statistic	IGT14 value	Desired range
Mean item difficulty	0.56	[0.3, 0.9]
Mean discrimination index	0.45	≥ 0.3
Mean point biserial coefficient	0.46	≥ 0.2

TABLE VII. The WLE and EAP reliability values [55], along with the infit and Rasch parameter ranges for the IGT14.

Rasch	IGT14 value
WLE reliability	0.680
EAP reliability	0.711
Infit	[1.13, 0.88]
Parameter	[2.10, −2.18]

referring to this version as IGT14. The psychometric properties of the 14-item version of the test can be found in the Appendix, see Tables VI–VIII and the values of the CFA in Fig. 9. As we did not conduct interviews with the students, we can only guess why they selected certain answers. The students performed the test under the supervision of their teacher, who was responsible for conducting the test after instruction by the research team. Furthermore, validity evidence was collected in only one specific educational context—namely, with students in their first or second year of upper secondary school physics. Future studies should investigate the applicability and validity of the inventory in other populations, such as university students.

VI. CONCLUSION

Understanding how graphs change under Galilean transformations is a crucial aspect of graph mastery. In this study, we presented a novel Inventory of Galilean Transformations of uniform linear motion in position-time graphs. The inventory comprises 15 items that assess 3 transformations between stationary and moving FoRs. The test was validated by experts in physics education research and educational psychology. The IGT exhibits a high degree of internal consistency, as evidenced by the mean

item difficulty, mean discrimination index, and mean point-biserial coefficient, all of which exceed the established thresholds. Factor analysis demonstrated that the three transformations manifest as factors within our model. The application of IRT further confirmed the ability of the IGT to discriminate between the three transformations. This evidence supports the use of the IGT as a valid and reliable instrument for evaluating students' understanding of Galilean transformations at the upper secondary school level. The IGT can be used as a formative assessment either at the beginning of the course to assess the students' prior knowledge or at the end of the course to assess their learning progress. As the test covers a subsection of kinematics, it can also be used as a summative assessment after the section about changing FoRs. Additionally, the IGT can serve as a standardized instrument for summative assessment that allows comparisons of learning gains between different groups of students.

ACKNOWLEDGMENTS

The authors would like to thank all the students and teachers who participated in the study, Martin Mohr and Roman Schmid for the test development, Barbara Gänz for the review of the data evaluation, and Peter Edelsbrunner for the statistical consultation.

DATA AVAILABILITY

The data that support the findings of this article are openly available [58].

APPENDIX: STATISTICAL VALUES OF IGT14

The appendix presents the statistical values for the IGT without the last item 15; we therefore call this version the IGT14. We recommend using the IGT14 or modifying item 15.

TABLE VIII. For each item of the IGT14, the table includes the mean value of the item, alpha if deleted, the Rasch parameter, infit and outfit statistics, point biserial, item difficulty, and the item discrimination index.

Items	Mean	Alpha if deleted	Rasch parameter	Outfit	Infit	Point biserial	Item difficulty	Item discrimination
1	0.84	0.69	−1.96	0.92	0.92	0.49	0.84	0.37
2	0.66	0.69	−0.81	1.02	1.01	0.48	0.66	0.53
3	0.86	0.69	−2.18	0.73	0.90	0.50	0.86	0.34
4	0.80	0.68	−1.70	0.79	0.88	0.57	0.80	0.49
5	0.75	0.69	−1.30	0.92	0.97	0.50	0.75	0.49
6	0.35	0.71	0.74	1.26	1.13	0.33	0.35	0.32
7	0.75	0.69	−1.34	0.96	1.00	0.46	0.75	0.43
8	0.73	0.69	−1.21	0.98	1.01	0.47	0.73	0.47
9	0.56	0.70	−0.29	1.07	1.05	0.45	0.56	0.50
10	0.39	0.69	0.55	1.02	0.98	0.51	0.39	0.59
11	0.27	0.70	1.19	1.11	1.03	0.41	0.27	0.42
12	0.43	0.68	0.33	0.96	0.95	0.55	0.43	0.67
13	0.32	0.70	0.93	1.14	1.02	0.43	0.32	0.48
14	0.15	0.71	2.10	1.37	1.07	0.27	0.15	0.22

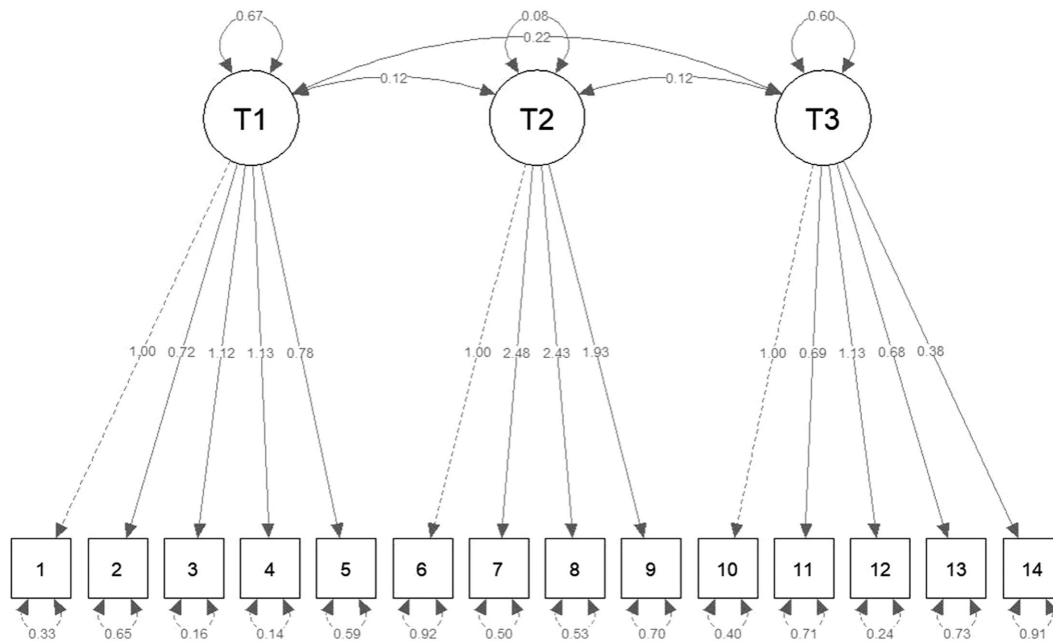


FIG. 9. CFA for the IGT14; the numbers correspond to the item number.

- [1] M. Caprile, R. Palmen, P. Sanz, and G. Dente, Encouraging STEM studies: labour market situation and comparison of practices targeted at young people in different Member States European Union, 2015, p. 38, [http://www.europarl.europa.eu/RegData/etudes/STUD/2015/542199/IPOL_STU\(2015\)542199_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/STUD/2015/542199/IPOL_STU(2015)542199_EN.pdf).
- [2] B. Cowie and B. Cooper, Exploring the challenge of developing student teacher data literacy, *Assess. Educ.* **24**, 147 (2017).
- [3] M. Planinic, L. Ivanjek, A. Susac, and Z. Milin-Sipus, Comparison of university students' understanding of graphs in different contexts, *Phys. Rev. ST Phys. Educ. Res.* **9**, 020103 (2013).
- [4] P. Klein, A. Lichtenberger, S. Küchemann, S. Becker, M. Kekule, J. Viiri, C. Baadte, A. Vaterlaus, and J. Kuhn, Visual attention while solving the test of understanding graphs in kinematics: An eye-tracking analysis, *Eur. J. Phys.* **41**, 025701 (2020).
- [5] P. Alstein, K. Krijtenburg-Lewerissa, and W.R. van Joolingen, Teaching and learning special relativity theory in secondary and lower undergraduate education: A literature review, *Phys. Rev. Phys. Educ. Res.* **17**, 023101 (2021).
- [6] S. Panse, J. Ramadas, and A. Kumar, Alternative conceptions in Galilean relativity: Frames of reference, *Int. J. Sci. Educ.* **16**, 63 (1994).
- [7] R. J. Beichner, Testing student interpretation of kinematics graphs, *Am. J. Phys.* **62**, 750 (1994).
- [8] A. Lichtenberger, C. Wagner, S.I. Hofer, E. Stern, and A. Vaterlaus, Validation and structural analysis of the kinematics concept test, *Phys. Rev. Phys. Educ. Res.* **13**, 010115 (2017).
- [9] P. Barniol and G. Zavala, Test of understanding of vectors: A reliable multiple-choice vector concept test, *Phys. Rev. ST Phys. Educ. Res.* **10**, 010121 (2014).
- [10] D. Hestenes, M. Wells, and G. Swackhamer, Force concept inventory, *Phys. Teach.* **30**, 141 (1992).
- [11] S. I. Hofer, R. Schumacher, and H. Rubin, The test of basic Mechanics Conceptual Understanding (bMCU): Using Rasch analysis to develop and evaluate an efficient multiple choice test on Newton's mechanics, *Int. J. STEM Educ.* **4**, 18 (2017).
- [12] S. Flegr, S. Becker, S. Kuechemann, K. Scheiter, and J. Kuhn, Development and validation of the ray optics in converging lenses concept inventory, *Phys. Rev. Phys. Educ. Res.* **18**, 020131 (2022).
- [13] P. Barniol and G. Zavala, Mechanical waves conceptual survey: Its modification and conversion to a standard multiple-choice test, *Phys. Rev. Phys. Educ. Res.* **12**, 010107 (2016).
- [14] J.S. Aslanides and C.M. Savage, Relativity concept inventory: Development, analysis, and results, *Phys. Rev. ST Phys. Educ. Res.* **9**, 010118 (2013).
- [15] E. Saltiel and J.L. Malgrange, "Spontaneous" ways of reasoning in elementary kinematics, *Eur. J. Phys.* **1**, 73 (1980).

- [16] Z. Tanel, Student difficulties in solving problems concerning special relativity and possible reasons for these difficulties, *J. Balt. Sci. Educ.* **13**, 573 (2014).
- [17] P. Klein, S. Gröber, J. Kuhn, A. Fleischhauer, and A. Müller, The right frame of reference makes it simple: An example of introductory mechanics supported by video analysis of motion, *Eur. J. Phys.* **36**, 015004 (2015).
- [18] M. Kozhevnikov, J. Gurlitt, and M. Kozhevnikov, Learning relative motion concepts in immersive and non-immersive virtual environments, *J. Sci. Educ. Technol.* **22**, 952 (2013).
- [19] J. Ramadas, S. Barve, and A. Kumar, Alternative conceptions in Galilean relativity: Inertial and non-inertial observers, *Int. J. Sci. Educ.* **18**, 615 (1996).
- [20] J. Larkin, J. McDermott, D. Simon, and H. Simon, Expert and novice performance in solving physics problems, *Science* **208**, 1335 (1980).
- [21] S. Brand-Gruwel, I. Wopereis, and Y. Vermetten, Information problem solving by experts and novices: Analysis of a complex cognitive skill, *Comput. Hum. Behav.* **21**, 487 (2005).
- [22] S. Klinaku, Galilean transformation in polar coordinates and Doppler effect, *Results Phys.* **31**, 104885 (2021).
- [23] L. Ding and R. Beichner, Approaches to data analysis of multiple-choice questions, *Phys. Rev. ST Phys. Educ. Res.* **5**, 020103 (2009).
- [24] D. G. Janelle, M. Hegarty, and N. S. Newcombe, Spatial thinking across the college curriculum: A report on a specialist meeting, *Spat. Cognit. Comput.* **14**, 124 (2014).
- [25] M. Kozhevnikov, M. A. Motes, and M. Hegarty, Spatial visualization in physics problem solving, *Cognit. Sci.* **31**, 549 (2007).
- [26] J. Ramadas, S. Barve, and A. Kumar, Alternative conceptions in Galilean relativity: Distance, time, energy and laws, *Int. J. Sci. Educ.* **18**, 463 (1996).
- [27] R. E. Scherr, P. S. Shaffer, and S. Vokos, Student understanding of time in special relativity: Simultaneity and reference frames, *Am. J. Phys.* **69**, S24 (2001).
- [28] L. C. McDermott, M. L. Rosenquist, and E. H. van Zee, Student difficulties in connecting graphs and physics: Examples from kinematics, *Am. J. Phys.* **55**, 503 (1987).
- [29] S. Wineburg, J. Breakstone, S. McGrew, M. D. Smith, and T. Ortega, Lateral reading on the open Internet: A district-wide field study in high school government classes, *J. Educ. Psychol.* **114**, 893 (2022).
- [30] J. E. Fan, Drawing to learn: How producing graphical representations enhances scientific thinking, *Transl. Issues Psychol. Sci.* **1**, 170 (2015).
- [31] E. B. Merki, S. I. Hofer, A. Vaterlaus, and A. Lichtenberger, Evaluating an interactive 360-degree learning environment for Galilean transformation of uniform motion in graphs developed on the basis of analyzed peer discussions, Doctoral thesis, ETH Zürich, 2025, [10.3929/ethz-b-000738033](https://ethz-b-000738033).
- [32] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/ww1x-kr12> for the English version IGT in PDF format.
- [33] R Core Team, *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria, 2024), <https://www.R-project.org/>.
- [34] A. Oosterhof, *Classroom Application of Educational Measurement* (Prentice Hall, Englewood Cliffs, NJ, 2001).
- [35] P. Potthoff, Measurement theory for the behavioral sciences: Edwin E. Ghiselli, J. P. Campbell, and S. Zedeck, San Francisco: W. H. Freeman and Company, 1981, pp. 491, \$27.95, *Math. Soc. Sci.* **4**, 319 (1983).
- [36] H. W. Marsh, K.-T. Hau, J. R. Balla, and D. Grayson, Is more ever too much? The number of indicators per factor in confirmatory factor analysis, *Multivar. Behav. Res.* **33**, 181 (1998).
- [37] Y. Rosseel, LAVAAN: An R package for structural equation modeling, *J. Stat. Software* **48**, 1 (2012).
- [38] P. M. Bentler and D. G. Bonett, Significance tests and goodness of fit in the analysis of covariance structures, *Psychol. Bull.* **88**, 588 (1980).
- [39] L. t. Hu and P. M. Bentler, Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives, *Struct. Equation Model.* **6**, 1 (1999).
- [40] R. E. Schumacker and R. G. Lomax, *A Beginner's Guide to Structural Equation Modeling*, 4th ed. (Routledge, New York, 2015), [10.4324/9781315749105](https://doi.org/10.4324/9781315749105).
- [41] J. H. Steiger, Structural model evaluation and modification: An interval estimation approach, *Multivar. Behav. Res.* **25**, 173 (1990).
- [42] S. Epskamp, semPlot: Unified visualizations of structural equation models, *Struct. Equation Model.* **22**, 474 (2015).
- [43] F. B. Baker, *The Basics of Item Response Theory*, 2nd ed. (ERIC Clearinghouse on Assessment and Evaluation, College Park, MD, 2001), <https://eric.ed.gov/?id=ED458219>.
- [44] G. Rasch, *Studies in Mathematical Psychology: I. Probabilistic Models for Some Intelligence and Attainment Tests* (Nielsen & Lydiche, Oxford, England, 1960).
- [45] B. Wright and M. M. C. Mok, An overview of the family of Rasch measurement models, in *Introduction to Rasch Measurement: Theory, Models and Applications*, edited by E. V. Smith, Jr., and R. M. Smith (JAM Press, Maple Grove, MN, 2004), pp. 1–24, <https://api.semanticscholar.org/CorpusID:17169559>.
- [46] J. Linacre, Sample size and item calibration stability, *Rasch Meas. Trans.* **7**, 328 (1994), <https://www.researchgate.net/publication/235361463>.
- [47] J. Linacre, Data variance: Explained, modeled and empirical, *Rasch Meas. Trans.* **17**, 942 (2003), <https://www.rasch.org/rmt/rmt173g.htm>.
- [48] R. F. Waugh and E. S. Chapman, An analysis of dimensionality using factor analysis (true-score theory) and Rasch measurement: What is the difference? Which method is better?, *J. Appl. Meas.* **6**, 80 (2005), <https://pubmed.ncbi.nlm.nih.gov/15701946/>.
- [49] A. Robitzsch, T. Kiefer, and M. Wu, TAM: Test analysis modules, R package version 4.2-21, <https://CRAN.R-project.org/package=TAM> (2024).
- [50] P. Mair and R. Hatzinger, Extended Rasch modeling: The eRm package for the application of IRT models in R, *J. Stat. Software* **20**, 1 (2007).
- [51] R. William, PSYCH: Procedures for psychological, psychometric, and personality research, version 2.4.6, <https://CRAN.R-project.org/package=psych> (2024).

- [52] H. Wickham, *GGPLOT2: Elegant Graphics for Data Analysis* (Springer-Verlag New York, 2016).
- [53] R. L. Doran, Basic measurement and evaluation of science instruction. National Science Teachers Association, 1742 Connecticut Ave., N.W., Washington, DC 20009 (Stock No. 471-14764; no price quoted), <https://files.eric.ed.gov/fulltext/ED196733.pdf>.
- [54] S. B. Green and Y. Yang, Reliability of summed item scores using structural equation modeling: An alternative to coefficient alpha, *Psychometrika* **74**, 155 (2009).
- [55] R. J. Adams, Reliability as a measurement design effect, *Stud. Educ. Eval.* **31**, 162 (2005).
- [56] B. D. Wright, Comparing Rasch measurement and factor analysis, *Struct. Equation Model.* **3**, 3 (1996).
- [57] A. Marzari, M. Di Mauro, T. Rosi, P. Onorato, and M. Malgieri, Investigating the principle of relativity and the principle of equivalence in classical mechanics: Design and evaluation of a teaching–learning sequence based on experiments and simulations, *Educ. Sci.* **13**, 712 (2023).
- [58] E. B. Merki, A. Lichtenberger, and S. I. Hofer, IGT Data (V3.0), OSF, [10.17605/OSF.IO/2WH4P](https://osf.io/2WH4P).