



Formative Assessment Strategies by Monitoring Science Students' Problem-Solving Skill Development

Frits F. B. Pals¹ · Jos L. J. Tolboom²  · Cor J. M. Suhre¹

Accepted: 7 November 2023
© The Author(s) 2024

Abstract To be able to support students' competence development in solving physics problems over the course of a lesson series effectively, teachers need a proper appreciation of students' deficiencies. As teachers commonly assess students' competence by means of written tests, teachers are challenged to interpret students' work on these tests and to intervene when some students fail to understand the proper application of solution methods in different contexts. This paper evaluates a formative assessment practice where teachers have been instructed to pinpoint students' level of understanding of kinematics problems by means of a cognitive diagnostic instrument and to provide personalized hints that match students' current level of understanding. The study is novel in this sense that the assessment of written tests results of students' problem solving is not expressed in grades and pass rates, but in terms of cognitive level of understanding. The results show that teachers can determine and monitor shifts of performance of students' cognitive level of understanding by using this instrument. Second, the results indicate that among students with low initial results, the group that received sufficient feedback via sticky notes made significantly more progress in solving problems than the group that did not receive feedback on sticky notes. Third, the timing of feedback in the form of sticky notes did not affect on students' progress in achieving mastery at the end of the instruction period. Our conclusion of this study is: Cognition develops through levels and tiers and support is essential to move to the Zone of Proximal Development. Subsequently, we evaluate group and subgroup implications for didactic interventions and propose suggestions for further investigations.

Résumé Pour pouvoir soutenir efficacement le développement des compétences des élèves en résolution de problèmes de physique sur un ensemble de leçons, les enseignants ont besoin d'une appréciation juste des lacunes des élèves. Comme les enseignants évaluent généralement les compétences des élèves au moyen de tests écrits, il devient difficile pour eux d'évaluer le travail des élèves lors de ces tests et d'intervenir quand certains de ces élèves ne parviennent pas à comprendre comment bien appliquer

✉ Jos L. J. Tolboom
j.tolboom@slo.nl

¹ University of Groningen, Groningen, The Netherlands

² SLO, Netherlands Institute for Curriculum Development, Stationsplein 1, 3818 LE Amersfoort, The Netherlands

des méthodes de résolution dans différents contextes. Dans cet article, nous examinons une pratique d'évaluation formative dans le cadre de laquelle les enseignants ont reçu comme directive de déterminer le niveau de compréhension des élèves à l'égard des problèmes de cinématique à l'aide d'un instrument de diagnostic cognitif et de fournir des indications personnalisées qui correspondent au niveau de compréhension actuelle des élèves. Les résultats indiquent qu'en utilisant cet instrument, les enseignants sont en mesure de constater et de suivre les écarts de performance en ce qui a trait au niveau de compréhension cognitive des élèves. Deuxièmement, on constate que, parmi les élèves qui ont obtenu de faibles résultats initiaux, le groupe qui a reçu une rétroaction appropriée par le biais de papillons adhésifs a beaucoup plus progressé dans la résolution de problèmes que le groupe qui n'a pas bénéficié de rétroaction par papillon adhésif. Troisièmement, le moment choisi pour la rétroaction sous forme de papillon adhésif n'a pas eu d'incidence sur la progression vers l'atteinte de la maîtrise des élèves à la fin de la période d'enseignement. De plus, on aborde ensuite les implications didactiques en ce qui a trait aux écarts de performance des niveaux de compréhension cognitive des élèves, les tactiques à adopter pour fournir des conseils d'orientation et les limites entourant cette étude.

Keywords Secondary science education · Formative assessment · Problem solving · Cognition · Diagnostic assessment · Mastery-oriented instruction

Introduction

To solve physics problems is a demanding challenge for science students, because it requires integration of knowledge about specific concepts and solution methods. Students therefore regularly need feedback to master problem solving in different physics topics. However, the common approach of teachers to provide students feedback by means of grades on summative tests often fails to promote students' reflection on their solution methods (Khan et al., 2020). Without receiving hints on incorrect parts of their solution methods, students may lose faith in being able to master solving science problems and turn away from mathematics, STEM-related subjects, or careers in technical occupations (Gottfried et al., 2013; Taconis & Kessels, 2009; Mangels et al., 2012; Li & Schoenfeld, 2019). There is therefore an urgent need to investigate how teaching practice in physics can be improved to inform science teachers by effective formative assessment methods to support students in their learning progress, with the focus on basic cognitive processes (Chi et al., 1981; McDermott & Larkin, 1978; Mayer & Wittrock, 2009; Pressley et al., 1996). In this paper, we evaluate a formative assessment practice that addresses the close alignments of hints with respect to the integration of concepts and solution methods of that part of students' solutions method that indicates lack of understanding.

Formative Assessment Cycle

Essential in any instruction period are three components of the formative assessment cycle: (1) setting the desired learning goals, (2) assessing students' level of mastery of these goals, and (3) feedback and feed forward with adequate support to achieve the set learning goals (Black & William 2009; Ramaprasad, 1983; Taras, 2010). Each of the elements of the formative assessment cycle is interconnected and forms a sequentially influencing active part of the cyclical triangle, as described by Van den Berg et al. (2018). Vardi (2013) has described the "dialogue feedback cycle" of Beaumont et al. (2011), and links the subsequent actions of teachers' written comments to "allowing" students to move to the next stage of learning—in accordance with the "Zone of Proximal Development" of Vygotsky (1986; Kozulin, 2012, p. xxi). The Zone of Proximal Development refers to a range of abilities a student can already perform under guidance of an instructor, but not yet on his own.

These theories view the mastery of skills as a context-dependent process consisting of a sequence in restructuring students' conceptions of ways to complete sets of tasks. The development of a skill requires students to understand connections between these conceptions to allow for integration of these insights in students' metacognition concerning the execution of the skill.

Setting the Desired Goals

This study is situated around the instruction of the goal/subject kinematics. To set the learning goals is then to clarify concepts of, for instance (instantaneous) speed, acceleration in combination with graphical representations and calculations with adequate units. Once the instructions have been given, students have to reach (actively) the desired goals in these lesson series with interaction of teachers and peers.

Assessment: The Need to Distinguish Student's Actual Level of Understanding

During the course of teaching with respect to setting the desired goals, teachers have to decide (Moon, 2005), at the start of physics and/or mathematics support (Anderson, 2007), what additional information students need in order to continue to master problem solving, for teachers assessing students' level of mastery of learning goals and providing appropriate feedback to students is not as easy as it seems (Alonzo, 2018; Amels-de Groot, 2021; Gotwals et al., 2015; Heritage et al., 2009; Khan et al., 2020; Schneider & Gowan, 2013; Tolboom, 2012; Van der Steen et al., 2019). To determine deficient parts of students' solution methods and being able to align feedback and further instructional support to support students' understanding of solution methods can be a huge challenge. A possible method to accurately determine students' current deficiencies in problem solving is described by the skill theory (Fischer, 2008; Fischer & Bidell, 2006) and dynamic systems theory (Van Geert & Fischer, 2009). These theories describe and suggest that support that aligns closely with students' current level of skill mastery provides a solid basis for contemplation of next steps for skill improvement.

In an earlier stage, we have developed a diagnostic instrument (Pals et al., 2023), based on principles laid out by Fischer (1980, 2008; De Bordes, 2013), and adapted this instrument for physics education. This instrument focuses on two kinds of levels of understanding in particular. First, the cognitive *functional* level of understanding students have reached (Fischer, 1980; Schwartz & Fischer, 2005; Pals et al., 2023). This means: Students can complete a task (or a part) without additional teacher assistance. Second, if students complete all (or a part) of a task *with* additional teacher (or peer) assistance, they reach the cognitive *optimal* level of understanding, the so-called Zone of Proximal Development (Fischer, 1980; Schwartz & Fischer, 2005; Vygotsky, 1978, p. 85).

The instrument was developed to allow teachers to determine students' functional level of understanding in mastery of problem solving in a specific physics area and to use this knowledge for providing students to reach the cognitively *optimal* level of understanding by personalized hints, with respect to the formative assessment cycle. The change between the functional and optimal level of understanding can give teachers insight how students develop over the series of lessons and can encourage students to overcome challenges and stumbling blocks (Galbraith & Stillman, 2006).

Pinpointing Students' Level of Understanding with the Instrument

To develop a solid integrated knowledge base, it is important that students link between accurate self-generated representations (such as reading and drawing, memories, and calculations). Monitoring changes of students' cognitive levels of understanding how to solve science problems, teachers need to evaluate systematically students' understanding of three basic sequential cognitive activities for correct physics problem solving (Table 1): sensorimotor level of understanding (a graphical representation of

Table 1 Code table developed by Fischer (1980) and De Bordes (2013) containing three tiers of level of understanding, with numbers and descriptions of observations

Level of understanding	<i>n</i>	Description of observation
Sensorimotor	1	SRs* that link to visual information of the task
Representational	2	SRs that link to visual information of the task to unobservable relations (e.g., thoughts, formulas)
Abstract	3	SRs that link to the formulation of mathematical expressions

SRs self-generated representations

the problem), representational level of understanding (applying right formulas), and abstract level of understanding (applying correct formulas).

After a first crude assessment of student's ability to deploy these three activities correctly, teachers can then refine the diagnosis by indicating the highest (i.e., functional) cognitive level of understanding in the eleven levels of understanding as starting point for responding (Table 2).

The Provision of Personalized Hints

The application of the diagnostic instrument provides teachers a possibility to pinpoint systematically students' functional level of understanding and deficiencies in level of understanding. In the current study, the pinpointing of the functional level of understanding is followed by the provision of personalized hints (reaching the optimal level of understanding), by means of written feedback on sticky notes, to improve students' mastery of reaching the desired goals and self-regulation in problem-solving skills (Bao & Koenig, 2019; Tolboom, 2012; Scholtz, 2007; Shute, 2008; Vardi, 2013). The personalized hints are teachers' feedback/feed forward given on sticky notes on what went wrong and how to proceed, and tailored to match each students' level of understanding. Learning with understanding is essential to empower students to solve problems they face in the future. According to Mayer (2014), teachers had to provide feedback specific to errors with content-specific next step information (feed forward).

In this study, we opted for a switching replications design. Although the possibilities to provide personalized hints in a classroom setting using such a design are limited, several options of oral, audio, or written feedback and/or feed forward are available (Mayer, 2014). To achieve the desired goals, we selected a written feedback approach. The reasons for selecting this approach were pragmatic. A written feedback approach allowed us to split classes into two groups on which teachers could provide classroom instruction to all students and alternatively additional personal written information to only one group following a written test to assess students' problem solving. A second reason was to provide teachers a qualitatively personal communication method, which had to be closely aligned with students' level of understanding. Third according to Moser (2020, p. 105), to give students the possibility in "contributing to their own cognitive development and classmates." These considerations lead to the following research questions.

Research Questions

This study focuses on two aspects of formative assessment. First, identification of the primary cause of students' failure: an incorrect representation of the problem, the selection of an incorrect formula, or executing incorrect calculations. The second aspect is the effectiveness of teachers' provision of targeted help to students by giving hints on sticky notes. These two focuses are articulated by the following questions:

Table 2 Code table developed by Fischer (1980) and De Bordes (2013), adapted to include 11 levels of insight (capability)

Level of insight	<i>n</i>	Description of observation
Sensorimotor action	1	Signs of delineation, e.g., marks, symbols, or drawings, which cannot be coded in terms of levels of insight
Sensorimotor mapping	2	SRs* visualizing information of the task without relation and/or delineation between objects, or with relation without causality
Sensorimotor system	3	SRs visualizing information of the task with two or more relations and/or delineation between objects or relations with causality
Representational action	4	SRs linking to not observable relations and/or a link between a SR and a formula
Representational mapping	5	SRs linking to unobservable relations and/or a link between a SR and a formula (e.g., $s=vt$)
Representational system	6	SRs indicating the understanding of the solution method
Single abstraction	7	Student shows understanding the regularity/pattern/choice of an adequate formula
Multiple abstraction	8	Student shows understanding in correct applying the formula
Abstract conducting	9	Student shows understanding in calculating a formula
Conducting	10	Student produces an answer
Abstract mapping	11	Student correctly relates the answer to the assignment context

SRs self-generated representations, *n* level of insight

1. Do students who receive classroom instruction and additional personalized hints on sticky notes show more progression in physics problem-solving skill development than students who receive classroom instruction only?
2. Does the timing of application of additional personalized hints affect on students' cognitive progress in achieving mastery at the end of the instruction period?
3. Are there features that may be of interest to teachers and researchers using this cognitive diagnostic instrument to increase the match between the form of support and the need in support of students who are at various stages of competent scientific problem solvers during the formative period?

Method

Treatment

According to Wiklund-Hörnqvist et al. (2014), repeated testing in formative assessment is a proven method to stimulate education, and to McDaniel et al. (2015) who concluded that a test-restudy-test sequence has a higher learning and retention rate compared to a sequence of 3 times of restudy or a series of 3 times testing the material. To address the research questions, we used a pre-test–post-test design (Abbasian, 2016; Vygotsky, 1986) with switching replication design: the Formative Assessment Model (FAM). The FAM (Table 3) covered a period of 16 lessons as part of an existing course of science lessons in four weeks in each of the grades and followed by a school test (Lesson 17).

Three tests were administered by applying the cognitive diagnostic instrument (Table 1) to determine students' level of understanding of two groups in this period.

Each of the two formative tests lasted for maximal 10 min. Test 3 was a part of the final test. Every sequential test contained a corresponding task content type. Teachers conducted tests in real classroom setting, determined students' written responses by means of the diagnostic instrument, and formulate additional personalized hints on sticky notes based on the observed cognitive level of understanding.

Test 1 is a pre-test without intervention. After Test 1 was accomplished, Group I received additional personalized hints on sticky notes as intervention. After Test 2, Group II received additional personalized hints on sticky notes as intervention. Both groups received feedback in daily routine classroom moments after both tests. The first two tests were not calculated toward their grade. The third test was a summative test and contained a similar assignment as in the two formative tests. The similarity between the tasks is reflected in the outlining of the context, a similar graph, and the same type of question about calculating the vehicle's acceleration at a point in time.

Table 3 Formative Assessment Model (FAM) consists two formative tests and a school test, Group I and Group II, and number of lessons

Lesson	1–7	8	9	10 ^a	11	12	13	14 ^a	15–16	17
Group I		Test 1		Sticky notes		Test 2				Test 3 (exam)
Group II		Test 1				Test 2		Sticky notes		Test 3 (exam)

^aBoth groups received class explanation

Instruments

The cognitive diagnostic instrument has a hierarchical structure divided in three categories: a sensorimotor, a representational, and an abstract category (Table 1). Those three categories are subdivided into eleven cognitive levels of understanding (Table 2). The instrument has a satisfactory reliability for assessing students' level of problem solving (estimated Krippendorff's alpha of .84) (Hayes & Krippendorff, 2007; Pals et al., 2023). In real-time education, the intervention was organized as additional personalized hints on sticky notes (Table 3).

Participants

All participants were students of a school in the northern rural part of the Netherlands. The students followed two streams of education: 26 students followed Senior General Secondary School and 56 students followed Pre-University Education. Two qualified and experienced teachers taught these students and were involved with the tests and integrated these tests in their lesson series—students knew the “learning goal,” according to the curriculum. In the Netherlands, the curricula have similar descriptions of the basic principles of kinematics (according to the level of the stream). Both teachers received individual training about the diagnostic instrument from the researcher in approximately half an hour. From a total of 82 students (Table 4) who completed three tests, 93 data were obtained. One class (11**) made two tasks.

The students were told that they participated in an investigation about how they solved science problems and students of each class were randomly assigned into two groups: Group I and Group II. Students of Group I received sticky notes and were allowed to use the sticky notes in Test 2 and both groups were allowed to use the sticky notes in Test 3. All students were allowed to use a formula book (Verkerk et al., 2004).

Data Collection

Students' mastery level of problem solving on the three test occasions was assessed by the teachers by means of the diagnostic assessment instrument and resulted in scores on a scale from 0 to 11. These scores were used to analyse the effects of using personalized hints on sticky notes as feedback/feed forward after test administration on students' test performance.

Examples of Diagnosing and Responding

Figure 1 (College voor Toetsen en Examens, HAVO, 2012; Pals et al., 2023) is an example of a realistic STEM-related assignment of physics problems by which we want to illustrate applications, results, and considerations of the diagnostic instrument (Bao & Koenig, 2019) to identify science students' problem solving in terms of cognitive level of understanding.

It shows the (v, t) diagram, depicting the velocity v as a function of the time t . The corresponding text of the assignment states: “Aircraft are regularly subjected to severe tests. An example of such a test is the rejected take-off test (RTO). During an RTO, an aircraft accelerates to the speed to take off. Then the brakes are applied as hard as possible.” In this study (not in the exam), students had to answer the question: “Determine the acceleration on time $t = 10$ s.”

The answer of student A (Group II) in Test 2 (Fig. 1) has been chosen at random and is used as an example of teacher's decision-making and the application of feedback/feed forward (for decision

Table 4 Grades, number of teachers, male and female science students in the subject kinematics (without force), mean and standard deviation of age, and types of assignments and tests

			Grade		
			10 ^a	10 ^b	11 ^b
Teacher			I	II	II
Gender	Male	40	14	14	12
	Female	42	12	14	16
Total		82	26	28	28
Age	M		15.8	14.96	16.35
	SD		0.90	0.45	0.75
			Type and number of assignments		
Test 1			a_t^c	a_t^c	v_a and v_{max}^d
Test 2			a_t	a_t	v_a and v_{max}
Test 3			a_t	a_t	v_a and v_{max}

^aSenior general secondary school; H

^bpre-university education; V

^c a_t assignment acceleration at specific time

^d v_a and v_{max} assignments of average speed and assignments of maximum speed

rules, see Pals et al., 2023). This includes student’s drawing and calculation (the two bowed descending lines in the diagram of student’s answer are the “attention” lines of the teacher).

This example shows the considerations teachers face: There are several cognitive levels of understanding of student A present to determine at the same time. Science teachers have to be emphatic in their contact, when they receive signals of lack of understanding. In daily classroom, many students experience “cognitive load” (Sweller et al., 2011). It is the teachers’ task to observe and to note students, who do not know how to convert appropriately the received information into action (for instance, in this task: to draw a tangent line). Tracking back the incorrect solutions of student A: First, student A must draw a tangent line at $t = 10$ s and not the line connecting points (0,0) and (10,140). Second, student A did not notice the differences of the units listed in the axes. And third, incorrect calculation with units. This student needs additional differential support to improve problem-solving skills to reach the desired goals.

As shown in Fig. 2, the teacher responded to at least three deficiencies in the student A’s problem-solving skill. Although level of understanding 2 can be coded, because of incorrect reading and drawing, the basis of the incorrectness of student A is the choice of an inappropriate physics concept. The instrument advises the teacher to code for level 4 as assessment to reach the desired goals. This example shows that teachers had to prioritize (to decide) at which students’ level of understanding the personalized hints had to be addressed (the “where”), in order to increase the match of form of support and students’ learning needs (the “close alignments”) to reach the desired goals.

In Test 1 (Lesson 8), the answer of student A was (too) coded as cognitive level of understanding 4, because of an inappropriately chosen concept (the functional level of understanding). Figure 1 shows student A’s answer of Test 2 in Lesson 12. In the previous eleven lessons, student A had not been able to internalize the information from class instructions needed to solve the problem by an appropriate concept (there is no difference between the functional level of understanding and the optimal level of understanding). After Test 2, this student received additional support and in Test 3 student A has reached the desired goal (the optimal level of understanding): the progression of student A’s cognitive level of understanding has changed from 4 to 11. This example (at random chosen) does show that this student understands what he had do to resolve these errors.

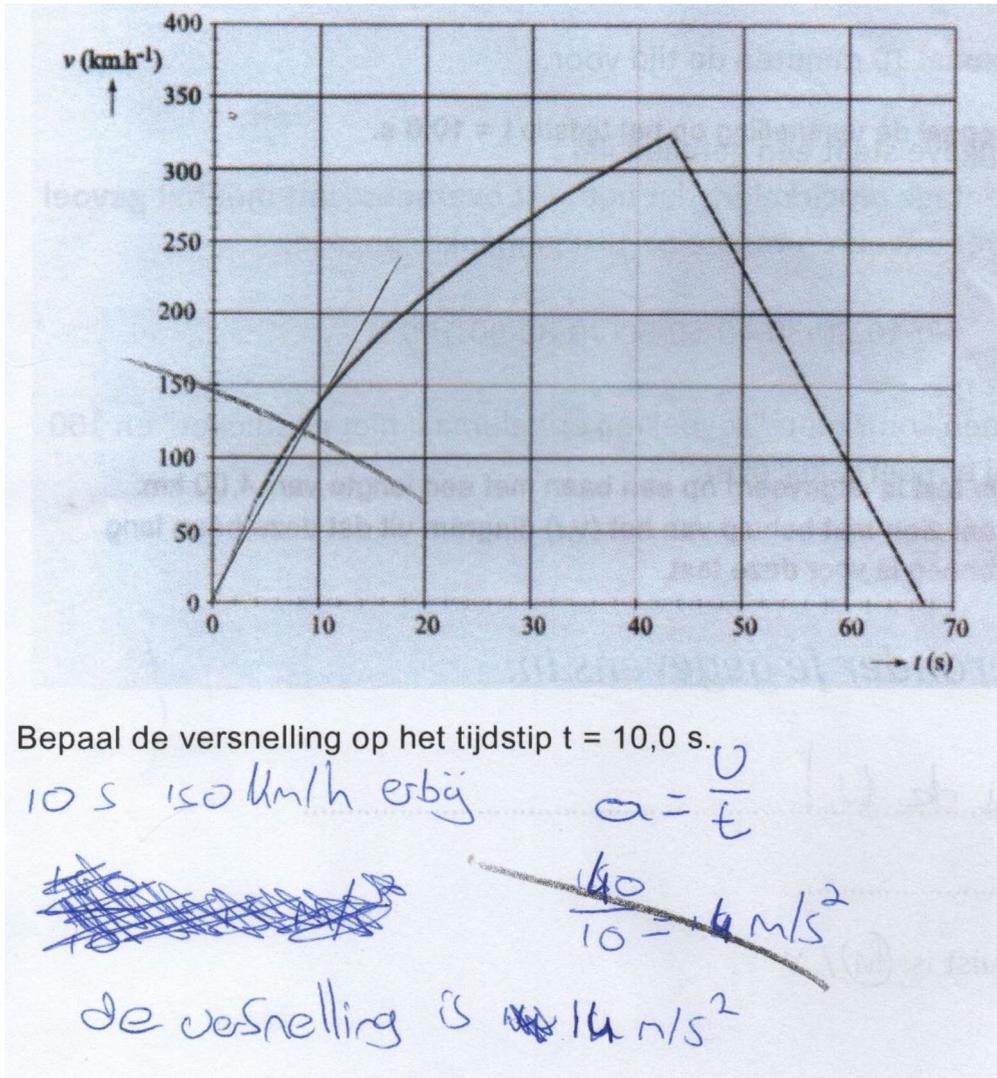


Fig. 1 The (v,t) diagram of an RTO test (College voor Toetsen en Examens, HAVO, 2012) administered of student A, in class 10**, Teacher II in Test 2, Lesson 12

Gain Scores and t -Tests

As shown above, student’s level of understanding of problem solving was determined by applying the cognitive diagnostic instrument on three tests (Table 1). For each student, differences in scores on subsequent occasions (gain scores) were computed to assess student’s progress in problem solving a physics task. To analyse the effectiveness of the personalized hints by means of sticky notes, we analysed the gain scores for both groups of students.

We ran separate analyses on the complete sample including students who already performed on a maximum level, and on a subsample of students who had not yet reached the maximum skill level (scores ≤ 10 on the assessment instrument). The reason for analysing this specific subset of students is that students who reach cognitive levels of understanding less than or equal to ten are the ones who clearly need additional support to improve their problem-solving skill (research questions 1 and 2).

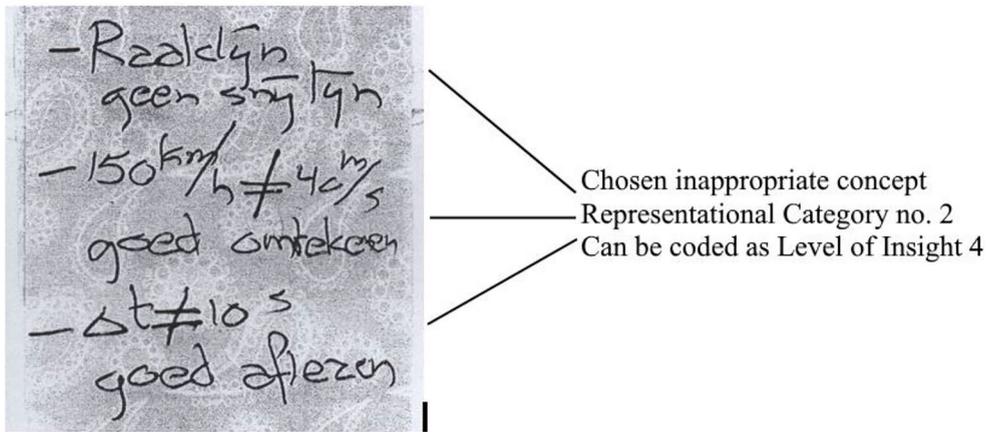


Fig. 2 An example of teacher’s responding of student A’s descriptive answer in Test 2 (Lesson 12, Group II) of the task mentioned in Fig. 3, subsequently researcher’s diagnosing and coding. *Note:* Translation of text: Tangent, no intersection; convert properly; read properly

We used independent *t*-tests to evaluate whether the scores of subtractions between subsequent test occasions differed significantly between groups.

Results

Results Between Groups in Three Tests

Table 5 displays the means and standard deviations of cognitive levels of understanding for Group I ($n=54$) and Group II ($n=54$) in three tests.

Table 5 shows an increase of the mean scores for both groups on T2 and T3. It turns out that the increase between Test 1 and Test 2 is higher in Group I compared to Group II, while the increase of Group II is higher between Test 2 and Test 3 compared to Group I. However, independent *t*-tests showed no significant differences in mean gain scores between both groups between Test 1 and Test 2 and between Test 2 and Test 3. But since a substantial number of students (37%) already performed at the maximum level of understanding at Test 1 and therefore could not achieve positive gain scores, we subsequently analysed the difference for the group of students who start at a level of understanding up to and including 10.

Table 5 Number of students of Group I and Group II, mean and standard deviation, and level of understanding of three tests, in eleven levels of understanding

	Test	T1	T2	T3
Group I ^a	Number	53	49	54
	Mean	6.53	8.18	8.96
	St dev	3,70	2.84	2.79
Group II ^b	Number	50	49	54
	Mean	6.62	7.42	9.04
	St dev	3.64	2.99	2.63

^aThe group received additional personalized hints on post-*t* notes after Test 1

^bThe group received additional personalized hints on post-*t* notes after Test 2

Table 6 Number of students of Group I and Group II, who started at level of understanding ≤ 10 , mean and standard deviation, in three tests

Test	Group I			Group II		
	1	2	3	1	2	3
<i>n</i>	30	30	30	30	30	30
Mean	4.29	8.25	8.32	4.2	6.63	8.4
St dev	2.25	2.81	3.13	2.13	2.97	2.87

Effects of Treatments on Students Starting at a Level of Understanding ≤ 10

Table 6 displays the means and standard deviations in eleven cognitive levels of understanding of students in Group I and Group II, who started in a level of understanding ≤ 10 (these students can make progression) in three tests. Logically, the means of the groups are lower than those in Table 5, especially for both groups in Test 1 (students whose result is level of understanding 11 will not be counted now).

For the group of students with scores from 0 up to and including 10, the independent one-tailed *t*-test revealed a significant difference *between* the two groups in mean gain scores between Test 1 and Test 2: $t(59)=2.03, p=0.02, r=0.52$. This result answers research question 1: “Do students who receive classroom instruction and additional personalized hints on sticky notes show more progression in physics problem-solving skill development than students who receive classroom instruction only?”. With one constraint: the proposition applies only to students with level of understanding 10 or below. Results show that the timing of the personalized hints has no significant effect on results between the two groups, eventually. This result answers research question 2: “Does the timing of application of additional personalized hints affect on students’ cognitive progress in achieving mastery at the end of the instruction period?”.

The differences between the two groups mentioned above can be visualized by means, additionally into box-plots. If students started at level of understanding ≤ 10 , the differences of students’ scores in T2 show that the median of Group I is at the same level as the third quartile of Group II.

An Analysis of Students’ Change of Cognitive Categories of Problem-Solving Skill Development

Figures 3 and 4 display the categorization of students’ understanding of problem-solving performance in the three main cognitive categories (sensomotoric, representational, and abstract) of both groups.

Figure 3 shows the increase of absolute numbers of students of Group I ($n=48$) in the *abstract* category (mathematics) in three tests: from 22 students in Test 1 to 40 students in Test 3 (respectively in percentages from 46 to 83%).

Fig. 3 Numbers of level of understandings of Group I ($n=48$) in three cognitive categories in three tests, in one series of 17 lessons

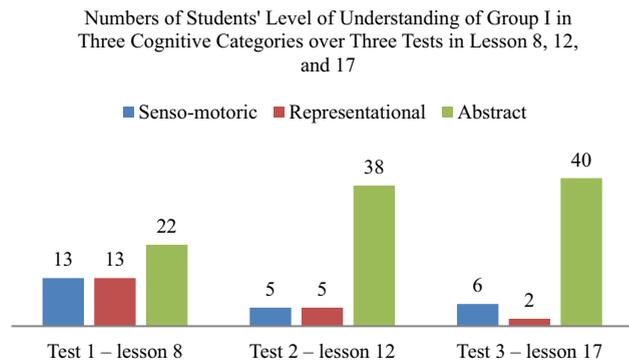


Fig. 4 Numbers of level of understandings of Group II ($n=45$) in three cognitive categories in three tests, in one series of 17 lessons

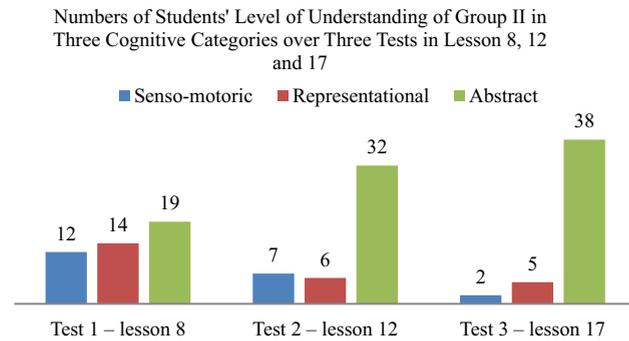


Figure 4 shows the numbers of students of Group II ($n=45$) increasing from 19 to 38 (respectively in percentages from 40 to 84%).

The frequencies show drops of numbers of students of both groups in the sensorimotor and representational categories (respectively), and the increase of numbers of students in the abstract category. Second, the numbers of students of both groups in the abstract category in Test 3 almost doubled compared to the numbers of students in the abstract category in Test 1. Third, comparing the observed levels of understanding as depicted in Figs. 3 and 4, it is noticeable that a number of students in Group 1 who still need help with drawing after three tests are higher than in Group 2.

Results show a total of 51 students of both groups have reached level of understanding 11 in Test 3 (55%): 27 students in Group I and 24 students in Group II. In Test 1, a total of 34 students started at level of understanding 11 (37%).

The *changes* of level of understanding of *individual* students are not (yet) noticeable. In the next section, the performances of cognitive levels of understanding of individual students in consecutive tests are determined and considered.

Identifying Differences of Subgroups of Students of Cognitive Level of Understanding and Consequences for Teachers' Personalized Support

By analysing randomly chosen examples of results of students in Group 1, we attempt to answer question 3: "Are there features that may be of interest to teachers and researchers using this diagnostic instrument in increasing the match between the form of support and students who are at various stages of competent scientific problem solvers during the formative period?"

Example 1: The ideal development? Figure 5 shows the progress in cognitive development of 12 students in Group I who started in the sensorimotor or representational category and concluded the lesson series at the (abstract) level of understanding 11.

Students who are determined (in the first test) in a level below level of understanding 7 have to be considered as students who need additional instruction.

Students in the sensorimotor category need specific support, i.e., to master correct reading and accurate drawing and need to receive suggestions to choose the correct concept to remember (because a well-executed reading/drawing can provide for instance a basis for a correct representation or executive calculation). This example shows one of the conclusions of this study: Students' cognition develops through levels of understanding and providing support is essential to move to the Zone of Proximal Development.

Students in the representational category need instruction to memorize and to connect the correct mathematical concept (formula) to see the problem and can be provided by suggestions about calculation. The students of both categories have to overcome many cognitive levels of understanding and may benefit from

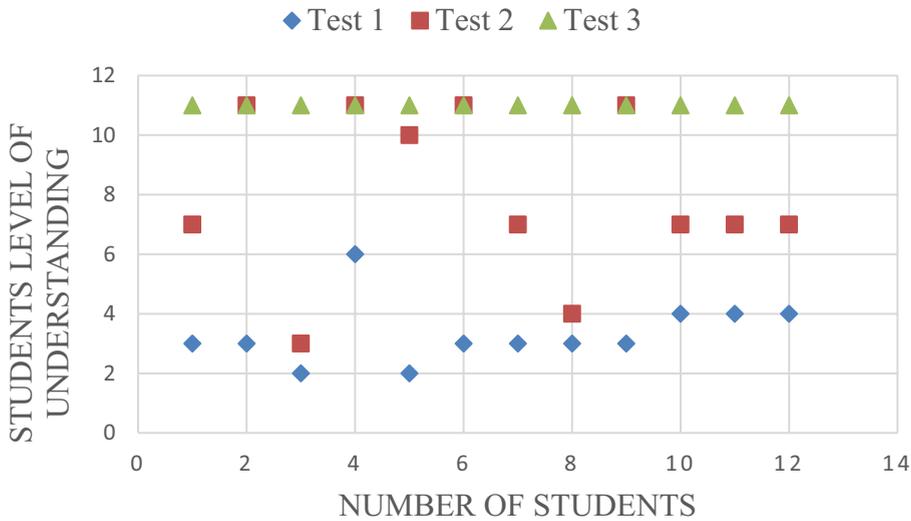


Fig. 5 The change of level of understanding of 12 students in Group I who ended in level of understanding 11 after three tests and started below level of understanding 7

targeted support/worked examples to grow and move into the abstract category and master problem-solving skills in a specific subject (Cooper & Sweller, 1987; Van Merriënboer & Sweller, 2005).

Example 2: Figure 6 shows 18 students in Group I who received the maximum score (level of understanding 11!) on Test 1, and shows changes to lower levels of understanding in the consecutive tests. These students need specific attention. We placed an exclamation point here, because teachers need to be aware of the decline in subsequent results of students who perform well at first glance.

Two-thirds of this group did not show the optimal level of understanding 11 in three tests, and five students (about one-third) were unable to stabilize at the level of understanding 11 in the final test.

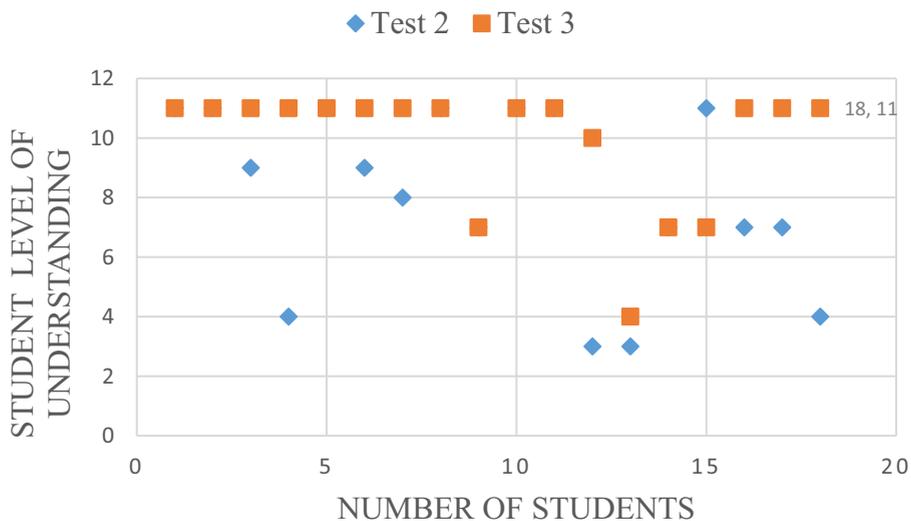


Fig. 6 The change of level of understanding in three tests of 18 students in Group I who started in level of understanding 11 in Test 1 (Test 1 is not included)

Although these students can be considered good students (they have started at level of understanding 11), the figure shows they still need close monitoring to prevent any relapse in a second (or third) test, as stated in the “Results” section. These students probably have built appropriate solution schemas to solve this type of science problems, and receive a positive result. To counteract, teachers have to appeal to their empathy to stimulate students’ motivation and conscientiousness to continue performing at level 11 of understanding to reach the desired goals. Citing Kalyuga et al. (2011, p. 29): “More expert learners already have such [incorporated of interacting elements, red.] schemas; thus, asking them to study the material is likely to constitute a redundant activity.” Teachers can pinpoint that building routine is now the emphasis for problem-solving skill and transfer (Cooper & Sweller, 1987).

Example 3: Figure 7 shows an erratic impression. This group of students of Group I started in cognitive level of understanding 4 in Test 1. Level of understanding 4 is the first level in the representation category and means—considering the instrument—students have chosen an inappropriate concept or an incorrect formula. It seems that this group can be divided into two groups. One part of this group is mentioned in Example 1 (increasing trend), but another (larger) part falls back into the sensorimotor category and/or was coded at a lower level of understanding in Test 3 than in Test 2.

This example shows another feature that teachers can look out for when providing specific support to students in this category. It seems that for students in Test 1 the cognitive level of understanding 4 can be considered as threshold: The personalized hints were sufficient as help in Test 2 (now, these students have reached the optimal level of understanding). But apparently, these hints have not been (fully) internalized for some students in Test 3.

This group of students (Fig. 7) can pose a real didactic challenge for teachers. These students seem to be unable to make connections between the information as described in the assignments and the prior knowledge they have of the instructions in preceding lessons. Even though this group received feed forward support on sticky notes after Test 1, more than six students ended on a lower level of understanding in Test 3 than in Test 2. Some students may escape from the teacher’s attention. A formative assessment method like the one presented here make this explicit, so the teacher can undertake action.

These three examples in kinematics (and the comparison of Figs. 3 and 4) show that students’ mastery of problem solving happens in leaps and bounds (cognitive and emotional) and is not predetermined as successive continuous increase of cognitive levels of understanding (Fischer, 2008; Schwartz and

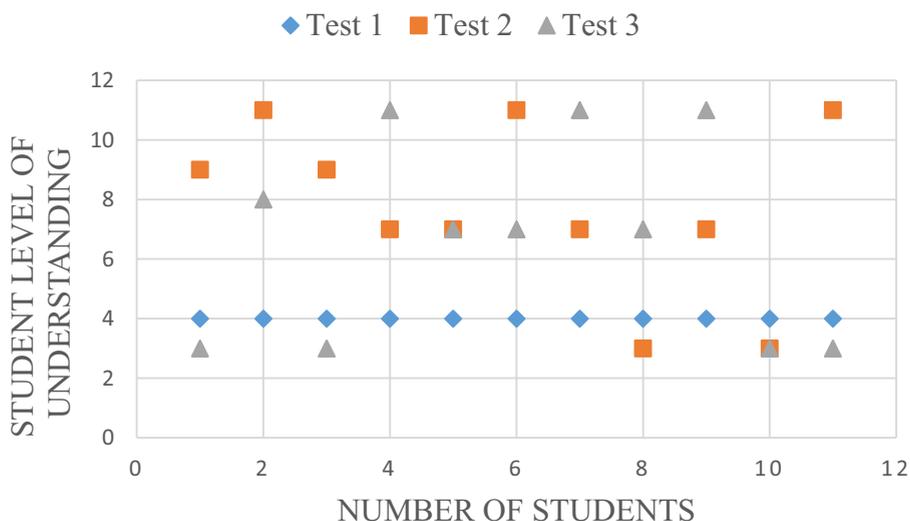


Fig. 7 The change of level of understanding in three tests of 11 students in Group I who started in level of understanding 4 in Test 1

Fischer, 2005; Van der Steen et al., 2019, pp. 6 and 7; Van Geert & Steenbeek, 2005, 2014) and have consequence for teachers' didactic actions and instructions. Investigations of reasons of backwards and forwards of results is beyond the scope of this study.

Discussion

Didactical Considerations

In this study, we explored the effectiveness of personalized support after formative tests by assessment of students' solution methods on tests by means of a cognitive diagnostic instrument. Teachers used this instrument to diagnose systematically the three core sets of cognitive categories of levels of understanding—sensorimotor, representational, and abstract category—which provide teachers a first indication of the students' cognitive level of understanding in applying formative assessment.

The first two categories (sensorimotor and representational) in student's orienting tasks should not be underestimated. According to Salleh and Zakaria (2009) and Norhatta et al. (2011), students tend not to take enough time to read and understand an assignment; reserving time and practicing patience ultimately have a positive effect on a thorough basis for the mathematical continuation of the task. Performing adequate mathematical operations is represented by the third cognitive (abstract) category.

By subdividing each of these three cognitive categories into a total of eleven levels of understanding, refinement of determination and monitoring can be achieved. The information obtained from determination can give teachers implications to provide students informed feedback/feed forward to solve problems, and educational adjustments. According to Bitchener and Storch (2016), could it give students something to hold on to if they know that science teachers are reviewing tests according to this procedure? First: Assessment of reading and drawing. Second: Assessment of remembering. And finally: Assessment of mathematics.

Applying the instrument and judicious assessment for close alignments can enhance the match of students' need to master problem solving (Shute, 2008; Vardi, 2013). At micro level, this diagnostic tool can provide teachers an accurate picture of how and where problems occurred in the subject matter, in class and individually. In other words: Provision of tailored feedback/feed forward and the decision-making skill can have an added value to whole classroom explanation of students' test performance. It would be interesting to investigate the relation between the decision-making skill of science teachers using this diagnostic instrument and the change of the professional subject-didactic capability of teachers.

When students receive their personalized hints (in this study on sticky notes), they have to motivate themselves to encounter a compliment or an error (sometimes several). And if it is an error, students must first recognize an error, acknowledge it, and finally erase the "old learning route" and apply and embed a "new learning route" as a pattern. These cognitive efforts can be: "What is the information - what do I need (what do I know or do not know yet) - how can I 'place' it (what do I need to know, to remember or to understand?) - and what should I do with it?" That is why it is important that a teacher discusses common mistakes afterwards with students, preferably at the end of a lesson so that all students may benefit from it. According to Kirschner et al. (2006, p. 77): "Learning is defined as a change in long-term memory." During physics lessons, students can experience excessive, confusing, and disappointing feelings while processing these cognitive efforts (Sweller et al., 2011; Hays et al., 2010). It could be interesting to investigate students' cognitive efforts by providing students with additional personalized hints on sticky notes and how students decide what the optimal benefit of information presented might be (Montague, 2002).

Two Sides of Cognitive Progress: Teaching and Learning

The personalized hints can be applicable as a tool for formative assessment for science teachers and researchers in monitored tests and lesson series, and can reveal students' progress in terms of cognitive level of understanding, from sensorimotor, representational to the abstract category, individually and as class. In addition to the three examples described, teachers can use the diagnosing and monitoring as reference how to group students and then to respond to this group in an adequate way to guide students or an individual student in the "Zone of Proximal Development." What is striking in the data is that the numbers of students in the abstract category in Test 1 (42%) are nearly the same as the numbers of students with level of understanding 10 or 11. This means (in this study) that half of the students (already) acquired an adequate solution method in the first formative test in Test 1. That may mean teachers can anticipate about two things at least: to try to stabilize half of this group at level of understanding 11 (see Example 2) and guide the other half of the students to change to a higher level of understanding (see Example 4). This can be a subject for further investigation: What does this dichotomy mean for science teacher's professionalism in dealing with educational demands in practice? Because guiding "weak and good" students with additional support has practical, organizational, didactical, and time management implications, and challenges to do justice to meet individual students' need (Amels-de Groot, 2021). We advocate that the teachers' attention at this point in students' learning process (see Examples 2 and 4) is crucial to avoid possible students' demotivation because of disappointments of one's own inability and as a result of experiencing this feeling (several times) the loss in interest in science. In general, a personalized dialogue can have added value, because then the possible cause of the lack in making content connections of an unexperienced learner, or an overload of information to the experienced learner can be traced to minimize "the expertise reversal effect" (Kalyuga et al., 2011). According to the conclusion of the authors: "To be efficient, instructional design should be tailored to the experience of the intended learners.", adequate contact between teachers and students enables novices to construct underpinned mental schemas, and can reduce cognitive load (Van Merriënboer & Sweller, 2005) and chronic stress (Tyng et al., 2017, p. 3). Teachers should be aware of the level of understanding of the second test of this group, because in the personalized contact after Test 2 they can assess whether student's level is stabilizing.

Another point of interest is the preservation of level of understanding. In education, it is common sense that students make progress in a period and increase their cognitive level of understanding by teacher instruction in daily class routine. This study shows that although personalized hints have a positive effect on students' progress on cognitive level of understanding, Fig. 7 (and the comparison of observed levels of understanding among students in the two research groups, depicted in Figs. 3 and 4) show profiles of whimsical performance shifts of student problem-solving skill development and suggest the impression that personalized hints and class instruction do not provide every student or group with adequate information to enhance problem-solving skills. Two questions for further research remain: How do students process information as written support in the lessons before a following test? And: How can teachers optimize the quality of personalized hints? In other words: How to reduce a possible mismatch in the quality of the information?

Although with research limitations (Van Geert & Steenbeek 2014, p. 34), this study endeavours to underline that formative assessment can provide science teachers with information and understanding about the cognitive development of students and what they can do with the information to provide feed forward and stimulate individually students' problem-solving skills, as complement of summative assessments (Scholtz, 2007).

Noteworthy is that results show that both groups' mean ended at level of understanding 8 (of students which started with levels below 11), and there is nearly no difference of the means of Group I between Test 2 and Test 3. This could have a reason in the differences of character and history between the subjects: physics and maths. The emphasis (of students) of math is about the calculation (level 8); the emphasis of physics (teachers) is the appropriate completion in terms of units and the use of significance

in physics (level 11). This means that these last levels have to be automatized too. Are math and physics really integrated in science education?

Conclusions

In this study, science teachers are challenged to provide timely appropriate personalized hints on sticky notes to improve students' competence in solving physics problems. By using the developed diagnostic instrument, teachers can monitor and evaluate students' deficiencies in problem-solving skills by determining their cognitive level of understanding (from level 0 up to including 11), and can adjust their formative assessment.

The emphasis of this study is to analyse data of groups of students with scores below a maximum score of level of understanding 10 (≤ 10). Applying the independent *t*-test to evaluate whether the (gain) scores of Group I, which received additional personalized hints on sticky notes, differed significantly from Group II in the subject of kinematics between subsequent test occasions Test 1 and Test 2.

This study shows that classroom instruction after a test can already make a difference in progress of students' cognitive level of understanding, and additional personalized hints on sticky notes increases this progression. Provision of tailor-made feed forward/feedback has an added value compared to solely explaining the test performance of students in class. Answering the first question: "Do students who receive classroom instruction and additional personalized hints on sticky notes show more who receive classroom instruction only?" The answer is yes, with one constraint: the proposition applies only to students with level of understanding 10 or below.

The answer of the second question: "Does the timing of the application of additional personalized hints affect on students' cognitive progress in achieving mastery at the end of the instruction period?" The answer is no. The timing of personalized support provides a booster in students' learning, but results show that it does not matter when the additional information is received eventually, as long as students receive the appropriate information that allow them to improve their problem solving timely before the final test (Van Merriënboer & Sweller, 2005). After two formative tests with similar physics tasks in 17 lessons compared with two classroom instructions and one group with additional personalized hints on sticky notes, there was no significant difference between the results of both groups in the final test.

To answer the third question: "Are there features that may be of interest to teachers (and researchers) using the diagnostic instrument in increasing the match between the form of support and the need in support of students who are at various stages of competent scientific problem solvers during the formative period?" Two features can be concluded: (1) Students might show several errors simultaneously in a task—teachers must prioritize in order to increase the match of form of support and students' learning needs. (2) Students' functional level of understanding can change (grow) to an optimal level of understanding (the level of understanding with support by a teacher). That does not mean that progressive cognitive change always will be stable over time. Guiding to the "Zone of Proximal Development" (Vygotsky, 1978) does not always lead to a stable higher cognitive level of understanding.

Two investigations can be recommended as formative assessment subjects for further research in two sides concerning the improvement of teaching and learning of cognitive progress and transfer of knowledge by using the diagnostic instrument: First, to set up a more intensive assessment teachers' practice to improve students' problem-solving skill, which is provided by Van den Berg et al. (2018). And second, to improve teachers' capacity to chance didactically, which is investigated by Amels-de Groot (2021).

The study shows that students' progression of cognitive development of solving specific physics problems can be measured in terms of levels of understanding and that the progression can be increased by personalized hints, which are closely aligned with the level of students' understanding. The results indicate that feed forward as formative assessment is an ongoing process of skills development support and is a challenge for teachers and students.

Declarations

Conflict of Interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abbasian, M. (2016). Dynamic Assessment: Review of Literature. *International Journal of Modern Language Teaching and Learning*, 1(3), 116–120.
- Alonzo A.C. (2018). An argument for formative assessment with science learning progressions. *Applied Measurement in Education*, 31(2), 104–112. <https://doi.org/10.1080/08957347.2017.1408630>
- Amels-de Groot, J. (2021). Teachers' capacity to realize educational change through inquiry-based working and distributed leadership. <https://doi.org/10.33612/diss.160149825>
- Anderson, T.R. (2007). Bridging the Gap; Bridging the Educational Research-Teaching Practice Gap; The power of assessment. *Biochemistry and molecular biology education*, 35(6), 471–477. <https://doi.org/10.1002/bambed.20135>
- Bao, L. & Koenig, K. (2019). Physics education research for 21st century learning. *Disciplinary and Interdisciplinary Science Education Research*, 1(2), 1–12. <https://doi.org/10.1186/s43031-019-0007-8>
- Beaumont, C., O'Doherty, M. & Shannon, L. (2011). Reconceptualising assessment feedback: A key to improving student learning? *Studies in Higher Education*, 36(6), 671–687. <https://doi.org/10.1080/03075071003731135>
- Bitchener, J., & Storch, N. (2016). *Written corrective feedback for L2 development*. Bristol: Multilingual Matters.
- Black, P., & William, D. (2009). Developing the theory of formative assessment. *Educational, Assessment, Evaluating and Accountability*, 21, 5–31. <https://doi.org/10.1007/s11092-008-9068-5>
- Chi, M.T.H., Feltovic, P.J. & Glaser, R. (1981). Categorization and Representation of Physics Problems by Experts and Novices. *Cognitive Science A Multidisciplinary Journal*, 5(2):121–152. https://doi.org/10.1207/s15516709cog0502_2
- College voor Toetsen en Examens. (2012, tijdvak 2). *Examenopgave HAVO, HAVO natuurkunde, opgave 2*. Stichting Studie Begeleiding.
- Cooper, G., & Sweller, J. (1987). The effects of schema acquisition and rule automation on mathematical problem-solving transfer. *Journal of Educational Psychology*, 79, 347–362.
- De Bordes, P. F. (2013). *Coderingsschema's non-verbaal gedrag*. Unpublished manuscript.
- Fischer, K.W. (1980). A theory of cognitive development: The control and construction of hierarchies of skills. *Psychological Review*, 87(6), 477–531.
- Fischer, K.W. (2008). Dynamic cycles of cognitive and brain development: Measuring growth in mind, brain, and education. In A.M. Battro, K.W. Fischer & P. Léna (Eds.), *The educated brain* (pp. 127–150). Cambridge University Press.
- Fischer, K. W., & Bidell, T. R. (2006). Dynamic development of action and thought. In W. Damon, & R. M. Lerner (Eds.), *Theoretical models of human development. Handbook of Child Psychology* (6th ed., pp. 313–399). Wiley.
- Galbraith, P.L., Stillman, G. (2006). A framework for identifying student blockages during transitions in the Modelling process. *Zentralblatt für Didaktik der Mathematik*, 38 (2), 143–162. <https://doi.org/10.1007/BF02655886>
- Gottfried, A.E., Marcoulides, G.A., Gottfried, A.W., Oliver, P.H. (2013). Longitudinal Pathways from Math Intrinsic Motivation and Achievement to Math Course Accomplishments and Educational Attainment. *Journal of Research on Educational Effectiveness*, 6(1), 68–92. <https://doi.org/10.1080/19345747.2012.698376>
- Gotwals, A.W., Philhower, J., Cisterna, D., Bennett, S. (2015). Using video to examine formative assessment practices as measures of expertise for mathematics and science teachers. *International Journal of Science and Mathematical Education*, 13, 405–423. <https://doi.org/10.1007/s10763-015-9623-8>

- Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, 1(1), 77–89.
- Hays, M.J., Kornell, N. & Bjork, R.A. (2010). The costs and benefits of providing feedback during learning. *Psychonomic Bulletin Review*, 17, 797–801. <https://doi.org/10.3758/PBR.17.6.797>
- Heritage, M.; Kim, J.; Vendlinski, T.; Herman, J. (2009). From Evidence to Action: A Seamless Process in Formative Assessment? *Educational Measurement: Issues and Practice*, 28(3), 24–31. <https://doi.org/10.1111/j.1745-3992.2009.00151.x>
- Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2011). The expertise reversal effect. *Educational Psychologist*, 38(1), 23–31. https://doi.org/10.1207/S15326985EP3801_4
- Khan, M., Zaman T.U., Saeed, A. (2020). Formative Assessment Practices of Physics Teachers in Pakistan. *Jurnal Pendidikan Fisika Indonesia*, 16 (2), 122–131. <https://doi.org/10.15294/jpfi.v16i2.25238>
- Kirschner, P., Speller, J., Clark, R.E. (2006). Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. *Educational Psychologist*, 41(2), 75–86. https://doi.org/10.1207/s15326985ep4102_1
- Li, Y. & Schoenfeld, A.H. (2019). Problematising teaching and learning mathematics as “given” in STEM education. *International Journal of STEM Education*, 6(44). <https://doi.org/10.1186/s40594-019-0197-9>
- McDermott, J., & Larkin, J.H. (1978). Re-representing textbook physics problems. Proceedings of the 2nd National Conference of the Canadian Society for Computational Studies of Intelligence. University of Toronto Press.
- Mangels, J.A., Good, C., Whiteman, R.C., Maniscalco, B., Carol S. Dweck, C.S. (2012). Emotion blocks the path to learning under stereotype threat. *Social Cognitive and Affective Neuroscience*, 7(2), 230–241. <https://doi.org/10.1093/scan/nsq100>
- Mayer, R. E. (2014). *The Cambridge handbook of multimedia learning* (2nd. ed.). New York, NY: Cambridge University Press.
- Mayer, R.E. & Wittrock, M. C. (2009). Problem solving. In A. P. Alexander, & P.H. Winne (Eds.), *Handbook of educational psychology*. (2nd ed., pp. 287–303). Taylor & Francis Group.
- McDaniel, M.A., Bugg, J.M. Liu, Y., Brick, J. (2015). When Does the Test-Study-Test Sequence Optimize Learning and Retention? *Journal of Experimental Psychology*, 21(4), 370–382. <https://doi.org/10.1037/xap0000063>
- Montague, M. (2002). Mathematical problem-solving instruction: Components, procedures, and materials. In M. Montague, & C. Warger (Eds.), *Afterschool extensions: Including students with disabilities in afterschool programs*.
- Moon, T.R. (2005). The Role of Assessment in Differentiation. *Theory in to Practice*, 44(3), 226–233. https://doi.org/10.1207/s15430421tip4403_7
- Moser, A. (2020). *Written Corrective Feedback: The Role of Learner Engagement: A Practical Approach*. Springer. Baden. Austria. <https://doi.org/10.1007/978-3-030-63994-5>
- Norhatta, M., Tengku, F.P., Tengku, M., Mohd, N.I. (2011). Factors that influence students in Mathematics achievements. *International Journal of Academic Research*, 3(3), 49–54.
- Pals, F.F.B., Tolboom, J.L.J., Suhre, C.J.M. (2023). Development of a formative assessment instrument to determine students’ need for corrective actions in physics: Identifying students’ functional level of understanding. *Thinking Skills and Creativity*, 50, 101387. <https://doi.org/10.1016/j.tsc.2023.101387>
- Pressley, M., Hogan, K., Wharton-McDonald, R., & Mistretta, J. (1996). The challenges of instructional scaffolding: The challenges of instruction that supports student thinking. *Learning Disabilities Research & Practice*, 11(3), 138–146.
- Ramaprasad, A. (1983). On the definition of feedback. *Systems Research and Behavioral Science*, 28(1), 4–13. <https://doi.org/10.1002/bs.3830280103>
- Salleh, F. & Zakaria, E. (2009). Non-routine problem solving and attitudes towards problem solving among high achievers. *The International Journal of Learning*, 16(5). <https://doi.org/10.18848/1447-9494/CGP/v16i05/46289>
- Schneider, M.C. & Gowan, P. (2013). Investigating Teachers’ Skills in Interpreting Evidence of Student Learning. *Applied Measurement in Education*, 26(3), 191–204. <https://doi.org/10.1080/08957347.2013.793185>
- Scholtz, A. (2007). An analysis of the impact of an authentic assessment strategy on student performance in a technology-mediated constructivist classroom: A study revisited. *International Journal of Education and Development using Information and Communication Technology*, 3(4), 42–53.
- Schwartz, M.S., & Fischer, K.W. (2005). Cognitive developmental change: Theories, models, and measurement. In A. Demetriou, & A. Raftopoulos (Eds.), *Building general knowledge and skill: Cognition and microdevelopment in science*. (pp. 157–185). Cambridge University Press.
- Shute, V.J. (2008). Focus on Formative Feedback. *Review of Educational Research*, 78(1), 153–189. <https://doi.org/10.3102/0034654307313795>
- Sweller, J., Ayers, P., & Kalyunga, S. (2011). *Cognitive load theory*. New York. Springer.
- Taconis, R., & Kessels, U. (2009). How choosing science depends on students’ individual fit to ‘science culture’. *International Journal of Science Education*, 31(8), 1115–1132. <https://doi.org/10.1080/09500690802050876>
- Taras, M. (2010). Assessment for learning: assessing the theory and evidence. *Procedia Social and Behavioural Sciences*, 2(2), 3015–3022. <https://doi.org/10.1016/j.sbspro.2010.03.457>
- Tyng, C.M., Amin, H.U., Saad, M.N.M., & Malik, A.S. (2017). The Influences of Emotion on Learning and Memory. *Frontiers in Psychology*, 8:1454. <https://doi.org/10.3389/fpsyg.2017.01454>
- Tolboom, J. L. J. (2012). *The potential of a classroom network to support teacher feedback: a study in statistics education*. <https://doi.org/10.33612/diss.14566050>

- Van den Berg, M., Bosker, R.J. & Suhre, C.J.M. (2018). Testing the effectiveness of classroom formative assessment in Dutch primary mathematics education. *School Effectiveness and School Improvement*. 29(3), 339–361. <https://doi.org/10.1080/09243453.2017.1406376>
- Van der Steen, S., Steenbeek, H.W., Den Hartigh, R.J.R., & van Geert, P.L.C. (2019). The Link between Microdevelopment and Long-Term Learning Trajectories in Science Learning. *Human Development*. 63(1), 4–32. <https://doi.org/10.1159/000501431>
- Van Geert, P., & Fischer, K.W. (2009). Dynamic systems and the quest for individual-based models of change and development. In J.P. Spencer, M.S.C. Thomas & J. McClelland (Eds.), *Toward a new grand theory of development? Connectionism and dynamic systems theory reconsidered*. Oxford University Press.
- Van Geert, P., & Steenbeek, H. (2005). The dynamics of scaffolding. *New Ideas in Psychology*. 23(3), 115–128. <https://doi.org/10.1016/j.newideapsych.2006.05.003>
- Van Geert, P., & Steenbeek, H. (2014). The good, the bad and the ugly? The dynamic interplay between educational practice, policy and research. *Complexity: An International Journal of Complexity and Education*. 1(2), 1–18.
- Van Merriënboer, J. J. G., & Sweller, J. (2005). Cognitive Load Theory and Complex Learning: Recent Developments and Future Directions. *Educational Psychology Review*, 17(2), 147–177. <https://doi.org/10.1007/s10648-005-3951-0>
- Vardi, I. (2013). Effectively feeding forward from one written assessment task to the next. *Assessment & Evaluation in Higher Education*. 38(5), 599–610. <https://doi.org/10.1080/02602938.2012.670197>
- Verkerk, G., Broens, J. B., Bouwens, R.E.A., Groot de, P.A.M., Kranendonk, W., Vogelesang, M.J., Westra, J.J., Wevers-Prijs, I.M. (2004). In NVON-commissie (Ed.), *Binas havo/vwo* (5th ed.). Wolters-Noordhoff.
- Vygotsky, L.S. (Ed.). (1978). *Mind and society: The development of higher mental processes*. Harvard University Press.
- Vygotsky, L.S. (1986). *Thought and language*. Newly revised and edited by A. Kozulin. (2012). MIT Press.
- Wiklund-Hörnqvist, C., Jonsson, B. & Nyberg, L. (2014). Strengthening concept learning by repeated testing. *Scandinavian Journal of Psychology*. 55(1), 10–16. <https://doi.org/10.1111/sjop.12093>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.