Discovering learning potential in secondary education using a dynamic screening instrument

Graafeiland, N. van; Veerbeek, J.; Janssen, B.; Vogelaar, B.

Article

Discovering Learning Potential in Secondary Education Using a Dynamic Screening Instrument

Nina van Graafeiland [1], Jochanan Veerbeek [2], Barbara Janssen [3] and Bart Vogelaar [4,*]

1   Department of Education, Utrecht University, 3584 CS Utrecht, The Netherlands
2   Institute of Education and Child Studies, Leiden University, 2333 AK Leiden, The Netherlands
3   Maris College, 2555 XZ Den Haag, The Netherlands
4   Institute of Psychology, Leiden University, 2333 AK Leiden, The Netherlands
*   Correspondence: b.vogelaar@fsw.leidenuniv.nl

Abstract: This study aimed to investigate the effectiveness of a newly developed dynamic screening instrument, using a learning phase with standardized prompts, to assess first year secondary school students' potential for learning. This instrument aimed to provide an alternative to the current static tests. The study sample included 55 children (mean age = 13.17 years) from different Dutch educational tracks. The dynamic screener consisted of the subtests reading, mathematics, working memory, planning, divergent thinking, and inductive reasoning. Each subtest employed a test–training–test design. Based on randomized blocking, half of the children received graduated prompt training between pre-test and post-test, while the other half did not. On some, but not all, subtests, training seemed to lead to an increase in performance. Additionally, some constructs measured through the dynamic screener were related to current school performance. This pilot study provides preliminary support for the use of such an instrument to gain more insight into children's learning potential and instructional needs. Directions for future research are discussed.

Keywords: dynamic testing; learning potential; secondary education; graduated prompts; instructional needs

1. Introduction

This paper concerns a pilot study, which aims to investigate the effectiveness of a newly developed dynamic screener to assess first year secondary school students' potential for learning. The dynamic screener aims to provide an alternative to the current static test procedures used to determine students' level in Dutch secondary education. To provide a broad overview of children's abilities, the screener taps into different potential predictors of academic success. By means of integrating feedback into the testing procedure, insight is gained into the child's progression in learning and the child's instructional needs, two frequently used measures to estimate potential for learning [1]. This information could potentially be used to provide an indication of suitable educational levels for students in the Dutch secondary school system. This study investigates whether the newly developed dynamic screener can be used to measure secondary school students' potential for learning in academic domains, executive functioning, and reasoning. In addition to expanding the current body of empirical studies on dynamic testing principles in a school context, the knowledge gained through this study can be used to adapt the ways in which children are placed into different educational tracks.

1.1. Transition from Primary to Secondary Education

After eight years of primary education, when Dutch children go to secondary school, they are subdivided into different educational tracks. Within these tracks, in general, differentiation is made between vocational and general education. Recommendations as to which track is suitable for an individual child are usually made on the basis of the primary

school end test (PSET) [2], in which children's current lingual and mathematical knowledge is assessed. Fifty percent of children are advised to enroll in one of the vocational education tracks, while the other fifty percent are directed towards one of the general education tracks. All tracks come with a different level and duration, and after successful completion, provide access to different forms of further education. An overview of the different tracks is provided in Table 1.

Table 1. Overview of Dutch secondary education tracks, ranked from lowest to highest level.

Track	Education Type	Duration	Future Educational Options
VMBO-B	Vocational	4 years	Basic professional college
VMBO-K	Vocational	4 years	Professional college
VMGO-G/T	Vocational	4 years	Specialist professional college
HAVO	General	5 years	University of applied sciences
VWO	General	6 years	University

Which track a child is allocated to therefore has a significant impact on their future. This leads to much critique, as this decision is already made when children are around the age of eleven or twelve years old [3]. In total, approximately 15% of students transfer to a higher or lower track during their time in secondary school [4]. Moreover, research has indicated that this percentage might be much higher, as under-recommended students are often unable to switch to a higher track due to self-fulfilling prophecy effects [5]. In contrast, over-recommendation is more favorable for children, because they seem to be generally capable of maintaining their optimistic track placement [6].

The decision of which track a child will be enrolled in is based on their results in the primary school end test and their teacher's advice. The combination of these two aspects is beneficial overall, as it leads to lower mis-recommendation rates, when compared to only using the PSET [7]. Using a test such as the PSET to estimate children's educational level has advantages, such as providing an objective insight into constructs that might be difficult for teachers to estimate for multiple individual students [7]. However, educational experts and practitioners often argue against the use of the PSET to advise students on which track they should enroll in. Among other things, these practitioners state that the PSET might enlarge the gap between children with different socio-economic backgrounds. As this test is perceived as very stressful for children, their parents and the primary schools, schools and parents are willing to spend a lot of energy, time, and money to prepare children for the PSET. As not all parents might have the opportunity to invest time and money into PSET preparation outside of the regular school curriculum, this leads to unequal opportunities for students [3]. Unequal opportunities are again reinforced by the language-dependent nature of the test. Students with lower lingual skills therefore score lower on the PSET than what might be expected based on their level of intelligence [8]. Additionally, the limited scope of the PSET is criticized, as the test is not able to measure latent abilities such as discipline, motivation, self-efficacy, creativity, and curiosity [9].

More importantly, the PSET is considered to be a static test. Such tests usually consist of a single-session format, in which children are required to solve tasks individually after standardized instruction [10]. This results in measuring prior developed knowledge and skills, rather than their latent abilities and their potential for learning [1]. The cognitive abilities of the child are therefore interpreted based on what they have learned in the past, and how this relates to the abilities of other children. A test which would enable measuring a child's performance before and after help would, as critics argue, provide a more balanced overview of their cognitive abilities. This overview indicates how much a child can learn, as their performance is compared to their own previous abilities, rather than the abilities of others [1].

1.2. Dynamic Testing

Therefore, as an alternative to static tests, the use of dynamic tests is advocated [1]. Dynamic testing builds on the concept of the zone of proximal development by Vygotsky [11]. This concept is defined as the difference between what someone can do without assistance, their zone of actual development, and what someone can do when receiving guidance, their zone of proximal development. Dynamic tests aim to tap into the zone of proximal development by providing a form of assessment in which instruction is integrated into the testing procedure. Instruction is often provided in the form of feedback or prompts, by which children are enabled to show progress in solving different sorts of cognitive tasks [12].

Dynamic tests exist in various forms and tap into different domains, but all of them include a form of feedback, hints, or instruction while participants solve the tasks [10]. An often-used design in dynamic testing is the test–training–test format, combined with a graduated prompt approach [13]. This approach entails help provided in the form of standardized prompts whenever a child fails to solve problems independently. These prompts are provided in a hierarchical manner, ranging from very general to very specific hints. General, metacognitive hints are provided, for instance, to activate children's prior knowledge. If this is not successful, cognitive hints tailored to each individual item can be used to help children solve these items. Lastly, modeling hints can be provided through a step-by-step explanation of the correct solution. Individual differences in need for instruction have been accurately measured through a graduated prompt approach in previous research [13]. In addition to providing insight into children's need for instruction, the test–training–test format provides insight into one's potential for learning by comparing pre-test performance, without the provision of feedback, to post-test performance, after the provision of feedback. Both performance change scores and instructional needs are frequently used measures of potential for learning [1]. However, the reliability of performance change scores is questionable, as double measurement error is incorporated into these scores [14]. Instructional needs, however, have been demonstrated to be valid measures of learning potential [15]. All in all, using both measures provides an extensive overview of individual differences in children's learning progress, and which types of instruction allow them to show this progress.

As yet, dynamic testing has not been used consistently to obtain information about children's potential school performance, despite evidence for the effectivity of curriculum-based dynamic tests. For example, batteries have been developed to dynamically assess reading comprehension [16,17], as well as mathematical problem solving [18,19]. Often, school performance is determined based on students' performance in static tests in several academic domains. Despite international differences in school systems, two domains are almost always deemed crucial: reading and mathematics [20]. When focusing on the Dutch secondary education system, these are two domains that are reflected in all the different tracks [21]. Both academic domains have been demonstrated to predict school achievement, and the extent to which children can adjust in the transition from primary to secondary education [22]. Most schools persistently use static measures to determine levels of reading and mathematical performance, despite evidence for the effectivity of dynamic tests to measure performance in these domains.

1.3. Executive Functioning

Another construct that plays a significant role in school performance is executive functioning. Executive functioning is defined as a mechanism by which performance is optimized in situations requiring the operation of a number of cognitive processes [23]. Executive functioning is an umbrella term that includes a variety of interrelated functions responsible for purposeful, goal-directed problem-solving behavior [24]. Research has indicated that executive functioning is an important predictor of school performance, potentially transcending IQ scores, and reading and mathematical skills [25]. This is possibly due to the fact that executive functioning is necessary to develop these reading [26]

and mathematical skills [27]. Additionally, successful learning, both inside and outside the classroom, requires students to complete tasks while retaining instructions and selectively ignoring distractions [28]. Finally, executive functioning has been shown to be necessary in the successful transition from primary to secondary education, as it is used to face academic and social challenges that are new to the child [29]. Recent research has found that executive functioning can be strengthened through intervention [30]. Dynamic tests have been applied successfully to various aspects of executive functioning, such as planning [31], working memory [32], and cognitive flexibility [14]. The tests used in these studies were all adapted from existing static measures for executive functioning, and we used a pre-test–training–post-test format to measure learning potential. In these studies, it was consistently found that testing executive functions dynamically provides sufficient insight into learning potential in these domains. The application of these dynamic tests could be beneficial to educational assessment.

1.4. Reasoning

Another main predictor of academic success is inductive reasoning, the cognitive process of comparing information or objects, and, if necessary, contrasting and transforming them in a novel manner [33]. This relationship is possibly explained by the need for a child to use inductive reasoning to apply their everyday learnings to new situations [34]. Therefore, students with better inductive reasoning skills are likely to perform better in school subjects such as mathematics, science, social studies, and languages [35]. Dynamic measures of inductive reasoning have been used before, successfully providing insight into the process occurring during the use of inductive reasoning [15,36]. These tests often make use of tasks including the solving of verbal or geometric analogies, seriation, or inclusion [15], and have shown that dynamic measures of inductive reasoning are related to school performance [37,38].

An additional aspect of reasoning involved in academic success seems to be divergent thinking [39,40]. This concept is defined as the capacity to generate creative ideas by combining diverse types of information in novel ways [41]. Similar to inductive reasoning, this process is involved in the application of everyday learning to new situations [42]. Research has indicated that divergent thinking and intelligence seem to be related from a young age [43]. Other studies have indicated divergent thinking to be a better predictor for academic success than intelligence measures [44,45], possibly due to the relationship between divergent thinking and extrinsic motivation. The application of dynamic testing to divergent thinking is still limited, but research has nevertheless demonstrated that it could be successful [46,47]. These studies have indicated that training during dynamic testing improves divergent thinking, and that dynamic measures can be used to identify different types of learning styles. These styles are identified by the different types of prompts children benefit from during training.

1.5. Current Study

Within this research project, a dynamic screener was developed, to possibly function as an alternative to the current static test used to classify children into the Dutch secondary education system, the PSET. For this purpose, it is important to obtain an overview of a child's potential for learning in different domains. Therefore, the newly developed screening instrument contains the following subtests: reading, mathematics, working memory, planning, divergent thinking, and inductive reasoning. These subtests were chosen because these constructs are key predictors of school performance [22,25,34,39].

The main goal of this study was to examine the validity and effectivity of the dynamic screener. In doing so, children's performance in the dynamic screener was examined and compared to their current school performance and advised educational level. Additionally, the psychometric properties of the instrument were investigated.

The first research question concerned the effect of training on the number of correctly performed items in all subtests. It was hypothesized that children who received this training

would show significantly more progress from pre- to post-test than children who were not trained, because dynamic testing allows children to show progress in different cognitive tasks [12].

Secondly, the relationship between all the measured constructs, advised educational level, and school performance was investigated. As this is an exploratory hypothesis, no specific relationships were expected.

Finally, children's instructional needs in each subtest were examined by looking at the average number of hints necessary to successfully complete an item during training. It was hypothesized that children in general education tracks would require a lower number of hints to improve their performance from pre- to post-test on all subtests. This expectation was formed due to the characterization of general education tracks by more independent learning, when compared to vocational education tracks [48]. Furthermore, the relationship between instructional needs and school performance was investigated. Again, this is an exploratory hypothesis, so no specific relationships were expected.

2. Materials and Methods

2.1. Design

This study had a pre-test–training–post-test experimental design. Children were allocated to one of two groups through a blocking procedure, based on educational level and gender. This procedure was performed to make sure any variability in results between the groups could not be attributed to these factors. Children in the experimental group received training between pre-test and post-test in all subtests. Children in the control group only participated in the pre-test and post-test. Table 2 displays an overview of the components of the dynamic screener administered under each condition.

Table 2. Overview of the study's design.

		Reading	Mathematics	Working Memory	Planning	Divergent Thinking	Inductive Reasoning
Control group	Pre-test	X	X	X	X	X	X
	Training						
	Post-test	X	X	X	X	X	X
Experimental group	Pre-test	X	X	X	X	X	X
	Training	X	X	X	X	X	X
	Post-test	X	X	X	X	X	X

2.2. Participants

Power analysis for repeated measures ANOVA was performed, with two groups and two measurement moments, to investigate the interaction of within- and between-subject factors. In this analysis, we aimed for a power of 0.95 and an effect size of 0.25, according to Cohen's recommendations [49]. This power analysis suggested the use of a sample of 55 participants. The final sample consisted of 55 students in the first year of secondary education (aged 12.35–14.36 years old, M = 13.17, SD = 0.51). All students were recruited at a single secondary school, with different locations of the school having different educational levels. The distribution of participants of different genders and educational levels under each condition is shown in Table 3.

2.3. Materials

2.3.1. Dynamic Screener

The dynamic screener consisted of 6 subtests, each concerning a specific domain. The domains that were tested through this tool were reading, mathematics, working memory, planning, divergent thinking, and inductive reasoning. The administration of each sub-test took about 30 min, leading to a maximum duration of 3 h. All items in the pre-test and post-test were matched to have a similar number of cognitive steps, and therefore, similar

difficulty levels. The subtests applied a graduated prompt approach during training; when an item was answered incorrectly, a metacognitive hint was first provided. This hint was followed by two more implicit, cognitive hints if the answer remained incorrect. If the answer was still incorrect after the first three hints, step-by-step modeling was used to explain the right answer. For some children, the test took longer than for others, because each subtest in the pre-test and post-test was ended after two consecutive wrong answers. This decision was made to avoid loss of motivation and concentration when children had to complete a lot of items that were outside of their abilities. Previous research applying dynamic testing has also used this stopping rule, e.g., [31]. Children were allowed to take 10 min breaks in between each subtest if necessary.

Table 3. Distribution of participants of different genders and educational levels under each condition.

		Vocational Tracks	General Tracks	Total
Control group	Boys	7	6	13
	Girls	9	5	14
Experimental group	Boys	10	7	17
	Girls	6	5	11
Total		32	23	55

Reading. This subtest used a pre-test–training–post-test design, in combination with a graduated prompt training approach [37]. Each session (pre-test, training, and post-test) consisted of two texts with different difficulty levels. These texts were derived from existing Dutch reading comprehension tests, developed by Nieuwsbegrip [50]. Each text was accompanied by 6 test-items. This led to the opportunity to discriminate between levels of students' skills in the different domains.

Mathematics. This subtest again used a pre-test–training–post-test design, in combination with a graduated prompt training approach [37]. Each session of this domain consisted of 16 items of varying difficulty levels. The 4 subdomains tested through this subtest were numbers and variables, ratios, geometry, and algebra. The items within these domains were derived from existing Dutch mathematical problems, developed by W4Kangoeroe [51].

Executive functioning. This domain consisted of two subtests: working memory and planning. The subtest for working memory was based on an existing dynamic test [32]. This subtest asked participants to recall picture sequences, again, in items with varying difficulty: three- to seven-item sequences. The subtest for planning was based on the dynamic version of the Tower of Hanoi [31], with 8 puzzles of varying difficulty levels in the pre-test, training, and post-test.

Divergent thinking. This subtest was based on the Guilford Alternative Uses Task [52]. The task asked participants to think of unusual ways to use different objects in a short time frame. The pre-test and post-test both consisted of 4 different items, with general training in divergent thinking in between. This training did not use a graduated prompt approach, but focused on teaching children new strategies in the use of divergent thinking.

Inductive reasoning. Lastly, this subtest was based on an existing dynamic test [53] consisting of 12 analogies of varying difficulty levels in the pre-test, training, and post-test.

2.3.2. Advised Educational Level

The educational level advised by the children's teacher at the end of primary school was used in this study, to compare them to the results of the dynamic screener.

2.3.3. School Performance

Recent school results for Dutch and mathematics from the school database were requested. We used the average grades for each subject and compare them to performance in the subtests of the dynamic screener. In the Netherlands, students receive grades on a scale from 1 to 10, with grades of 5.5 and higher being sufficient.

2.4. Procedure

Prior to data collection, approval from the Psychology Research Ethics Committee was obtained. Before participation in this study, informed consent was obtained from parents through Qualtrics survey software. Additionally, all participating children provided written informed consent. Only children who were able to provide consent from their parents and themselves were allowed to participate in this study. Data collection took place in school during regular school hours. A pen-and-paper version of the dynamic screener was administered through one-on-one interaction between a tester and a child.

2.5. Statistical Analyses

The first research question was investigated through a repeated measures multivariate analysis of variance (RM MANOVA). In this analysis, the session (pre-test vs. post-test) was included as a within-subjects factor. Condition (training vs. no training) was included as a between-subjects factor. The number of correct items in the sub-tests of mathematics, reading, divergent thinking, inductive reasoning, working memory, and planning were included as dependent variables.

To examine the second research question, correlation analyses were performed. First, Pearson correlations were calculated between the average school results for Dutch and mathematics and the pre-test scores in the subtests mathematics, reading, divergent thinking, inductive reasoning, working memory, and planning. This analysis was repeated on the post-test scores, separately for the control and experimental groups. Further, Spearman correlations were calculated between advised educational level and pre-test scores in the subtests mathematics, reading, divergent thinking, inductive reasoning, working memory, and planning. Again, the analysis was repeated on the post-test scores, separately for the control and experimental groups.

Lastly, the third research question was examined through the performance of one-way MANOVA. The average number of hints in mathematics, reading, working memory, planning, and inductive reasoning were included as dependent variables. Educational level functioned as the independent variable. Finally, Pearson correlations were calculated between instructional needs and average school results for Dutch and mathematics, on an exploratory level.

3. Results

3.1. Initial Analyses

First, the psychometric properties of the dynamic screener subtests were analyzed. To assess the difficulty level of each item, *p*-values were calculated by looking at the proportion of participants that solved the items accurately. Along with these *p*-values, item-total correlations were calculated. These scores reflect the relationship between individual items and the total subscale. Large variation in both the *p*-values and item-total correlations were found. When investigating these values, one should take into account that each item has a different sample base. This is because a stopping rule was applied after two consecutive wrong answers, and not all children completed all subtests. The full range of *p*-values and item-total correlations and the corresponding sample base can be found in Appendix A.

Cronbach's alphas were calculated for each domain in the pre-test to investigate the internal consistency of the dynamic screener. This analysis revealed moderate to good internal consistency for the subtests reading ($\alpha = 0.83$), mathematics ($\alpha = 0.84$), working memory ($\alpha = 0.50$), planning ($\alpha = 0.64$), divergent thinking ($\alpha = 0.67$), and inductive reasoning ($\alpha = 0.90$).

To investigate the influence of demographic factors on pre-test performance, multiple analyses were performed. No significant effects were found for SES, gender, and ethnicity. However, age was a significant predictor for reading performance, indicating that younger children scored higher in this subtest than older children.

3.2. Training Effects

The effect of training on performance in the dynamic screener subtests of all participants was examined through a repeated measures multivariate analysis of variance (RM MANOVA). The between-subject effects are shown in Table 4. The multivariate and univariate results are displayed in Table 5. The multivariate results indicate a significant effect of Session ($\lambda = 0.45$, $F(6,40) = 8.11$, $p < 0.001$, $\eta_p^2 = 0.55$), but not of Session x Condition ($\lambda = 0.83$, $F(6,40) = 1.34$, $p = 0.262$, $\eta_p^2 = 0.17$). On the univariate level, a significant Session effect was found for the subtests mathematics ($F(1,45) = 5.73$, $p = 0.21$, $\eta_p^2 = 0.11$), planning ($F(1,45) = 16.41$, $p < 0.001$, $\eta_p^2 = 0.27$), and divergent thinking ($F(1,45) = 4.13$, $p = 0.048$, $\eta_p^2 = 0.08$). No significant Session x Condition effects were found. Basic statistics for the different subtest scores are provided in Table 6. Additionally, the mean scores of all the dynamic screener subtests are displayed in Figure 1.

Table 4. Between-subject effects.

	F	*p*	η_p^2
Reading	0.001	0.979	<0.001
Mathematics	0.05	0.827	0.001
Working memory	0.88	0.354	0.02
Planning	3.35	0.074	0.07
Divergent thinking	0.08	0.776	0.002
Inductive reasoning	0.06	0.805	0.001

Table 5. RM MANOVA outcomes.

		λ	*F*	*p*	η_p^2
Multivariate	Session	0.45	8.11	<0.001	0.55
	Session × Condition	0.83	1.34	0.262	0.17
Reading	Session		0.51	0.479	0.01
	Session × Condition		3.15	0.083	0.07
Mathematics	Session		5.73	0.021	0.11
	Session × Condition		1.20	0.164	0.04
Working memory	Session		1.50	0.228	0.03
	Session × Condition		1.50	0.228	0.03
Planning	Session		16.41	<0.001	0.27
	Session × Condition		0.01	0.941	<0.001
Divergent thinking	Session		4.13	0.048	0.08
	Session × Condition		0.38	0.541	0.01
Inductive reasoning	Session		40.67	<0.001	0.48
	Session × Condition		3.37	0.073	0.07

Table 6. Basic statistics for scores of all dynamic screener subtests in pre- and post-test.

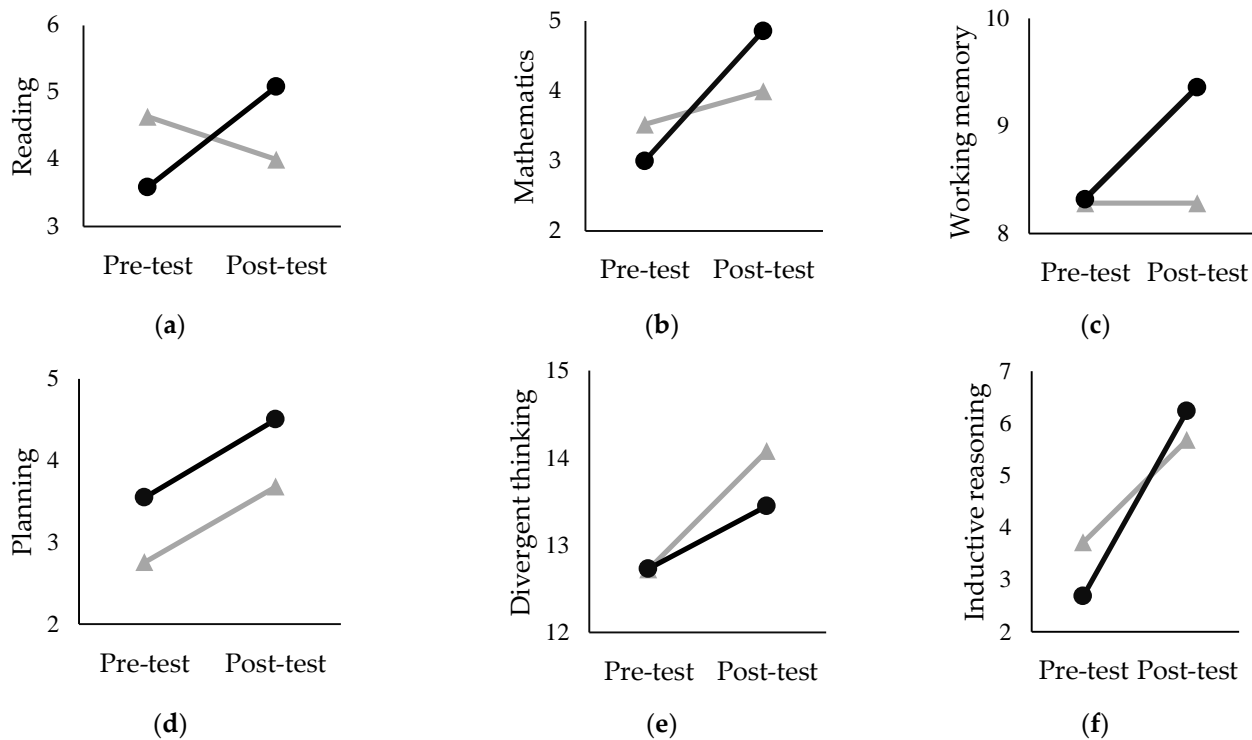|		Control Group		Experimental Group	
		Pre-Test	**Post-Test**	**Pre-Test**	**Post-Test**
Reading	*M (SD)*	4.64 (3.09)	4.00 (3.86)	3.59 (2.77)	5.09 (3.69)
Mathematics	*M (SD)*	3.52 (3.23)	4.00 (2.83)	3.00 (2.78)	4.86 (3.75)
Working memory	*M (SD)*	8.28 (2.41)	8.28 (2.42)	8.32 (2.40)	9.36 (2.84)
Planning	*M (SD)*	2.76 (1.74)	3.68 (1.82)	3.55 (1.57)	4.50 (1.63)
Divergent thinking	*M (SD)*	12.72 (3.53)	14.08 (3.27)	12.72 (4.57)	13.45 (4.94)
Inductive reasoning	*M (SD)*	3.72 (3.61)	5.68 (4.17)	2.68 (3.14)	6.23 (3.70)

Figure 1. Mean scores of dynamic subtests (**a**) reading, (**b**) mathematics, (**c**) working memory, (**d**) planning, (**e**) divergent thinking, and (**f**) inductive reasoning. The grey lines represent the control group, while the black lines represent the experimental group.

3.3. The Relationship between Dynamic Screener Scores and School Performance

Pearson correlations were calculated between the average school results for Dutch and mathematics and performance in the subtests reading, mathematics, working memory, planning, divergent thinking, and inductive reasoning in the pre-test and post-test. These correlations are displayed in Table 7.

Table 7. Correlations between school performance and dynamic screener scores.

| | **Pre-Test** | | **Post-Test** | | | |
| | | | **Control Group** | | **Experimental Group** | |
	Dutch	**Mathematics**	**Dutch**	**Mathematics**	**Dutch**	**Mathematics**
Reading	0.10	0.16	0.27	0.22	0.05	0.11
Mathematics	0.31 *	0.36 *	0.27	0.39 *	−0.15	0.01
Working memory	0.10	0.03	0.29	0.29	0.19	0.16
Planning	−0.08	0.06	−0.21	−0.16	0.06	0.35
Divergent thinking	0.23	0.16	0.19	<0.001	0.21	−0.09
Inductive reasoning	0.06	0.01	0.37	0.09	−0.13	0.04

* $p < 0.05$.

Further, Spearman correlations were calculated between advised educational level and performance in the subtests reading, mathematics, working memory, planning, divergent thinking, and inductive reasoning in the pre-test and post-test. These correlations are displayed in Table 8. A general pattern can be recognized; the dynamic test scores (in the experimental group) have a stronger relationship with the teacher's advice compared to the static test scores (in the control group).

Table 8. Spearman correlations between educational level and dynamic screener scores.

|	Pre-Test	Post-Test	
		Control Group	Experimental Group
Reading	0.19	0.33	0.51 *
Mathematics	0.38 *	0.63 *	0.64 *
Working memory	0.32 *	−0.01	0.40 *
Planning	0.19	0.05	0.65 *
Divergent thinking	0.07	−0.10	0.13
Inductive reasoning	0.44 *	0.35	0.69 *

* $p < 0.05$.

3.4. Instructional Needs

To investigate the differences in instructional needs between children with varying educational levels, another MANOVA was performed. The average number of hints in mathematics, reading, working memory, planning, and inductive reasoning were included as dependent variables. Educational level functioned as the independent variable. The results are displayed in Table 9. Additionally, a graphic overview of group means is available in Figure 2.

Table 9. MANOVA results for differences in instructional needs between educational levels.

	F	p	η_p^2
Reading	0.31	0.587	0.02
Mathematics	11.55	0.003	0.37
Working memory	0.06	0.803	0.003
Planning	2.52	0.128	0.11
Inductive reasoning	6.63	0.018	0.25
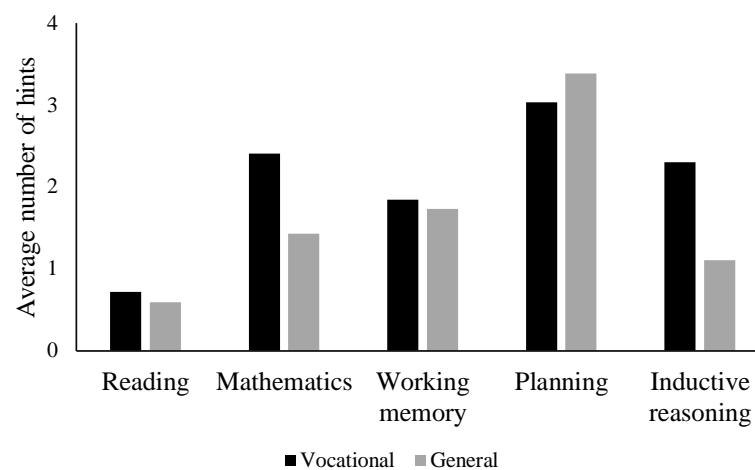
Figure 2. Group means of instructional needs.

Significant group differences were found for the subtest mathematics ($F(1,20) = 11.55$, $p = 0.003$, $\eta_p^2 = 0.37$). In fact, children in vocational education received more hints ($M = 2.40$, $SD = 0.76$) during training in this subtest than children in general education ($M = 1.43$, $SD = 0.49$). The same pattern was found for the subtest inductive reasoning ($F(1,20) = 6.63$, $p = 0.018$, $\eta_p^2 = 0.25$). Again, children in vocational education received more hints ($M = 2.30$, $SD = 1.04$) during training in this subtest than children in general education ($M = 1.81$, $SD = 1.21$).

Additionally, correlations were calculated between instructional needs and school performance in the subjects Dutch and mathematics. These correlations are displayed in Table 10, and no significant relationships were found.

Table 10. Correlations between school performance and instructional needs.

	Dutch	Mathematics
Reading	−0.35	0.15
Mathematics	0.09	−0.19
Working memory	0.22	−0.08
Planning	−0.25	−0.29
Inductive reasoning	−0.06	−0.12

4. Discussion

This study investigated whether our newly developed dynamic screener can be used to measure secondary school students' potential for learning in academic domains, executive functioning, and reasoning. The study provides preliminary support for the effectiveness of this tool in discovering learning potential in secondary education students, but resulted in mixed findings.

All groups of children showed progression from pre-test to post-test in the subtests mathematics, planning, divergent thinking, and inductive reasoning. This indicates the occurrence of a learning effect. Contrary to our hypotheses, trained children did not demonstrate larger improvements in performance in any of the subtests when compared to the control children. These findings could indicate that training might not have been effective. However, a detailed investigation of the mean scores suggests different learning slopes for children under different conditions in the subtests reading, mathematics, working memory, and inductive reasoning. This is in line with previous research that provides support for the use of dynamic testing to gain insight in to these cognitive abilities [15,18,19,32,36]. For the subtest reading, a decrease in performance from pre-test to post-test is observed in the control group, while the experimental group displays an increase in performance. This could be explained by a decrease in motivation, and/or concentration when reading multiple texts without intervening training. Performance in this subtest seems to have no relationship with performance in the school subjects Dutch and mathematics. Additionally, performance in the subtest reading in the pre-test does not relate to advised educational level. However, in the post-test, a significant relationship is found for the experimental group. This could indicate that dynamic testing is more in line with teachers' didactic expertise and expectations than static testing. Moreover, the control group shows little growth in performance from pre-test to post-test in the subtests mathematics, working memory, and inductive reasoning, while the experimental group shows a slight increase. These subtests all seem to have a relationship with advised educational level, but only mathematical performance relates to performance in the school subjects of both Dutch and mathematics. Although seemingly unexpected, the relationship between mathematical performance in the screener and the average grade for Dutch can be explained by the finding that solving mathematical problems initially requires some form of linguistic knowledge [54]. As expected, the relationship with the average grade for mathematics remains in the post-test under the control condition, but not in the experimental condition; dynamic testing is expected to relate less to static measures (e.g., how school performance is currently measured) than tests that do not employ this feedback-incorporated format. No relationship is found between school performance and the subtests planning and divergent thinking. Additionally, no group differences can be interpreted for these domains. This lack of findings could be explained by the difficulty of measuring these constructs [55,56]. Pen-and-paper administration of the dynamic screener increased the variability in instruction and scoring between testers. This could also provide an explanation for the finding of lower internal consistency in these subtests. Future research should include an investigation into ways to improve the measurement of these domains.

Regarding instructional needs, group differences for educational level were found in the subtests mathematics and inductive reasoning. These differences indicate that in general, children in general education can thrive from only receiving a few hints, while children in vocational education need a bit more help to obtain the right solution. This

was in line with our expectations based on the characterization of general education tracks by more independent learning, when compared to vocational education tracks [48]. However, this relationship was not found for any of the other subtests. Therefore, these subtests provide less insight into group differences in instructional needs. However, as was reflected in the standard deviations, large individual differences were found. Therefore, further research should also focus on the nature of the hints (metacognitive, cognitive, or modeling) [15] that different children benefit from.

Limitations

This study failed to provide significant results for most analyses, though investigating the mean scores could lead to the interpretation of expected patterns. However, it is important to note that any conclusion based on non-significant effects is purely speculation. While one should always be careful when interpreting insignificant models, the higher *p*-values in this study could be explained by our small sample size, and therefore, a lack of power. Future studies into the use of dynamic screening to discover secondary school students' learning potential should therefore aim to recruit a larger number of participants, preferably within different populations, as well. As this was a pilot study, the generalizability of the results is was gained through replication in different populations. This study was conducted within a single school, located in a large city in the Netherlands. Therefore, replication of this research, for example, in more rural areas, might lead to different results.

Due to the lower sample size, our design was limited to the use of one control group to uphold statistical power. As a consequence of this design, improvements found in the group of trained children might be explained by the simple fact that they completed more items between pre-test and post-test, and therefore, displayed a learning effect, separate from the effectivity of the training. However, previous research including two control conditions has demonstrated that improvements in trained children can actually be attributed to the effectivity of the training [57]. Future studies with larger sample sizes, preferably at least 100 participants per condition, should therefore focus on the inclusion of a second control condition.

Lastly, the inconsistency in our results might be explained by the pen-and-paper administration of the dynamic screener. Even though a standardized protocol was used, slight differences in the observed learning potential might be due to individual tester factors. This confounding factor can be eliminated in future research by using computerized dynamic testing with digitalized prompts. Computerized dynamic testing has been suggested to result in larger learning gains [1,58], and is also less time-intensive [59]. Additionally, computerized testing has been shown to have a positive effect on motivation [60], especially when incorporating game mechanics into these tests [61].

This study was a pilot study that investigated a newly developed instrument. While most results were non-significant, the inspection of found patterns provide a lot of information to further adapt and investigate this instrument. In addition to the aforementioned aspects, future research could focus on the roles of motivation and concentration in performance in the dynamic screener, and how to diminish this effect. For example, the item difficulty could be adjusted, as could the length of the training and the standardization of its protocol. Finally, the real promise of dynamic testing lies in the investigation of inter-individual differences in potential for learning. This was outside of the scope of the current study, but in future research utilizing this dynamic screening instrument, group and inter-individual patterns found could be studied in more detail.

5. Conclusions

In conclusion, the current study provides preliminary support for the use of our dynamic screener to gain more insight into secondary school students' learning potential. However, adaptation to fit the screener to the target group is necessary. Future research using different, larger sample sizes should provide more solid support for our findings.

After adaptation and validation, the screener could offer insights relevant to the way in which children are placed into different educational tracks. These insights include the children's potential for learning across different cognitive aspects, specifically mathematical performance, working memory, and inductive reasoning, and the types of instructions they benefit from. Investigating inter-individual differences in children's potential for learning and instructional needs could give a more accurate estimation of their educational level, leading to a decrease in children being placed on the wrong track. This could lead to an increase in overall wellbeing in these students, as well as broader options for their future education. Additionally, the adaptation of education to individual needs could be made a lot easier for schools and teachers with these insights. An instrument such as this could provide reports on children's performance, including learning potential scores and indications of their instructional needs on different constructs. This could make it easier for teachers to create subgroups of children who need different kinds of instructions and assignments to learn in class.

Appendix A

Table A1. *p*-values and item-total correlations for subtest reading.

	p-Values Accuracy			Item-Total Correlations			Sample Base		
	Pre-Test	Post-Test		Pre-test	Post-Test		Pre-Test	Post-Test	
Item		Control Group	Experimental Group		Control Group	Experimental Group		Control Group	Experimental Group
1	0.57	0.58	0.64	0.27	0.59	−0.001	51	26	25
2	0.61	0.43	0.52	0.27	−0.03	0.03	51	26	25
3	0.63	0.50	0.68	0.53	0.55	0.29	43	19	24
4	0.41	0.50	0.48	0.40	0.71	0.33	41	17	20
5	0.67	0.35	0.48	0.53	0.50	0.16	36	15	17
6	0.45	0.50	0.44	0.40	0.55	0.24	35	15	14
7	0.08	0.39	0.28	0.29	0.58	−0.12	35	14	13
8	0.33	0.50	0.48	0.43	0.55	0.26	27	13	12
9	0.06	0.39	0.32	−0.03	0.38	−0.02	18	13	12
10	0.22	0.08	0.16	0.50	0.10	−0.03	16	13	12
11	0.14	0.08	0.12	0.17	0.01	0.09	14	10	8
12	0.06	0.12	0.20	0.06	0.22	0.06	13	4	6

Table A2. *p*-values and item-total correlations for subtest mathematics.

	p-Values Accuracy			Item-Total Correlations			Sample Base		
	Pre-Test	Post-Test		Pre-Test	Post-Test		Pre-Test	Post-Test	
Item		Control Group	Experimental Group		Control Group	Experimental Group		Control Group	Experimental Group
1	0.57	0.89	0.79	0.24	−0.07	0.32	54	26	28
2	0.67	0.81	0.64	0.13	0.08	0.18	54	26	28
3	0.48	0.58	0.43	0.45	0.08	0.04	46	25	26
4	0.44	0.69	0.46	0.15	0.32	0.17	42	24	20
5	0.28	0.39	0.15	0.27	0.22	0.25	35	22	17
6	0.28	0.43	0.36	0.16	0.40	0.29	27	19	14
7	0.07	0.15	0.25	0.04	0.22	0.29	19	16	12
8	0.11	0.31	0.25	0.02	0.35	0.16	16	14	11
9	0.15	0.08	0.29	0.07	−0.20	0.25	9	9	10
10	0.07	0.08	0.14	0.11	−0.20	0.28	8	8	9
11	0.11	0.00	0.18	0.07		0.15	8	2	8
12	0.06	0.04	0.14	−0.02	−0.05	0.40	7	2	7
13	0.04	0.00	0.04	−0.01		−0.22	6	1	5
14	0.02	0.00	0.11	−0.01		0.58	4	1	4
15	0.00	0.00	0.11			0.38	3	0	4
16	0.00	0.00	0.07			0.17	1	0	4

Table A3. *p*-values and item-total correlations for subtest working memory.

	p-Values Accuracy			Item-Total Correlations			Sample Base		
	Pre-Test	Post-Test		Pre-Test	Post-Test		Pre-Test	Post-Test	
Item		Control Group	Experimental Group		Control Group	Experimental Group		Control Group	Experimental Group
1	0.97	0.96	0.96	0.22	0.36	0.09	55	27	28
2	0.95	0.96	0.89	0.33	0.52	0.02	55	27	28
3	0.75	0.89	0.89	0.09	0.61	0.20	55	27	28
4	0.69	0.85	0.68	0.23	0.21	−0.01	55	26	28
5	0.22	0.26	0.25	0.21	0.17	0.42	50	26	28
6	0.18	0.19	0.18	0.25	0.04	−0.05	44	24	24
7	0.06	0.11	0.04	−0.18	0.02	0.06	30	14	15
8	0.04	0.07	0.04	0.22	−0.15	0.06	19	9	7
9	0.00	0.00	0.04			0.06	9	3	1
10	0.00	0.04	0.00		−0.02		4	2	1

Table A4. *p*-values and item-total correlations for subtest planning.

	p-Values Accuracy			Item-Total Correlations			Sample Base		
	Pre-Test	Post-Test		Pre-Test	Post-Test		Pre-Test	Post-Test	
Item		Control Group	Experimental Group		Control Group	Experimental Group		Control Group	Experimental Group
1	0.96	1	1	−0.01			49	26	23
2	0.67	0.77	0.74	0.39	0.13	−0.21	49	26	23
3	0.65	0.89	0.83	0.33	0.50	−0.17	48	26	23
4	0.39	0.62	0.61	0.42	0.26	−0.07	37	24	22
5	0.25	0.54	0.30	0.50	0.38	0.28	35	23	20
6	0.14	0.39	0.26	0.47	0.52	0.39	23	18	14
7	0.06	0.15	0.09	0.35	0.24	0.31	14	15	8
8	0.00	0.04	0.09		−0.20	0.31	10	10	6

Table A5. *p*-values and item-total correlations for subtest inductive reasoning.

	p-Values Accuracy			Item-Total Correlations			Sample Base		
	Pre-Test	Post-Test		Pre-Test	Post-Test		Pre-Test	Post-Test	
Item		Control Group	Experimental Group		Control Group	Experimental Group		Control Group	Experimental Group
1	0.73	0.71	0.88	0.34	0.52	0.45	48	24	24
2	0.44	0.75	0.83	0.35	0.41	−0.18	48	24	24
3	0.44	0.67	0.71	0.42	0.89	0.12	39	19	23
4	0.31	0.71	0.71	0.55	0.61	−0.01	33	19	21
5	0.31	0.50	0.42	0.55	0.36	0.31	31	18	21
6	0.17	0.58	0.63	0.45	0.70	0.13	25	17	19
7	0.17	0.50	0.33	0.57	0.60	0.39	23	16	16
8	0.17	0.54	0.38	0.53	0.57	0.49	21	16	16
9	0.10	0.38	0.33	0.54	0.47	0.59	20	13	11
10	0.13	0.13	0.13	0.55	0.58	0.51	18	13	10
11	0.10	0.25	0.38	0.64	0.20	0.52	17	9	9
12	0.13	0.21	0.29	0.55	0.30	0.43	16	8	9

References

1. Resing, W.C.M.; Elliott, J.G.; Vogelaar, B. *Assessing Potential for Learning in School Children*; Oxford University Press: Oxford, UK, 2020.
2. Centrale Eindtoets PO. Available online: https://www.centraleeindtoetspo.nl/publicaties/publicaties/2015/05/01/toetswijzer-algemeen-deel-voor-eindtoets-po (accessed on 16 November 2022).
3. Elffers, L. *De Bijlesgeneratie: Opkomst van de Onderwijscompetitie*; Amsterdam University Press: Amsterdam, The Netherlands, 2018.
4. van Rooijen, M.; Korpershoek, H.; Vugteveen, J.; Opdenakker, M.C. De overgang van het basis- naar het voortgezet onderwijs en de verdere schoolloopbaan. *Pedagog. Stud.* **2017**, *94*, 110–134.
5. Jussim, L.; Harber, K.D. Teacher expectations and self-fulfilling prophecies: Knowns and unknowns, resolved and unresolved controversies. *Pers. Soc. Psychol. Rev.* **2005**, *9*, 131–155. [CrossRef] [PubMed]
6. Timmermans, A.; Kuyper, H.; van der Werf, G. *Schooladviezen en Onderwijsloopbanen. Voorkomen, Risicofactoren en Gevolgen van Onder- en Overadvisering*; Gronings Instituut voor Onderzoek van Onderwijs: Groningen, The Netherlands, 2013.
7. Lek, K. Teacher Knows Best? On the (Dis)Advantages of Teacher Judgments and Test Results, and How to Optimally Combine Them. Ph.D. Thesis, Utrecht University, Utrecht, The Netherlands, 2020.
8. Berends, I.; van Lieshout, E.C.D.M. The effect of illustrations in arithmetic problem-solving: Effects of increased cognitive load. *Learn. Instr.* **2009**, *19*, 3. [CrossRef]
9. Wij-Leren.nl. Available online: https://wij-leren.nl/10-pittige-problemen-met-de-centrale-eindtoets.php (accessed on 18 November 2022).
10. Sternberg, R.J.; Grigorenko, E. *Dynamic Testing*; Cambridge University Press: New York, NY, USA, 2002.
11. Vygotsky, L.S. *Thought and Language*; MIT Press: Cambridge, UK, 1962.
12. Elliott, J.G.; Grigorenko, E.; Resing, W.C.M. Dynamic assessment: The need for a dynamic approach. In *International Encyclopedia of Education*; Peterson, P., Baker, E., McGaw, B., Eds.; Elsevier: Oxford, UK, 2010; Volume 3, pp. 220–225.
13. Resing, W.C.M.; Elliott, J.G. Dynamic testing with tangible electronics: Measuring children's change in strategy use with a series completion task. *Br. J. Educ. Psychol.* **2011**, *81*, 579–605. [CrossRef]
14. Weingartz, S.; Wiedl, K.H.; Watzke, S. Dynamic assessment of executive functioning. (How) can we measure change? *J. Cogn. Educ. Psychol.* **2008**, *7*, 368–387. [CrossRef]
15. Resing, W.C.M. Dynamic testing and individualized instruction: Helpful in cognitive education? *J. Cogn. Educ. Psychol.* **2013**, *12*, 81–95. [CrossRef]
16. Gruhn, S.; Segers, E.; Keuning, J.; Verhoeven, L. Profiling children's reading comprehension: A dynamic approach. *Learn. Individ. Differ.* **2020**, *82*, 101923. [CrossRef]
17. Yang, Y.; Qian, D.D. Promoting L2 English learners' reading proficiency through computerized dynamic assessment. *Comput. Assist. Lang. Learn.* **2020**, *33*, 628–652. [CrossRef]
18. Fuchs, L.S.; Compton, D.L.; Fuchs, D.; Hollenbeck, K.N.; Hamlett, C.L.; Seethaler, P.M. Two-stage screening for math-problem-solving difficulty using dynamic assessment of algebraic learning. *J. Learn. Disabil.* **2011**, *44*, 372–380. [CrossRef]
19. Jeltova, I.; Birney, D.; Fredine, N.; Jarvin, L.; Sternberg, R.J.; Grigorenko, E. Making instruction and assessment responsive to diverse students' progress: Group-administered dynamic assessment in teaching mathematics. *J. Learn. Disabil.* **2011**, *44*, 381–395. [CrossRef]
20. Duncan, G.J.; Dowsett, C.J.; Claessens, A.; Magnuson, K.; Huston, A.C.; Klebanov, P.; Pagani, L.S.; Feinstein, L.; Engel, M.; Brooks-Gunn, J.; et al. School readiness and later achievement. *Dev. Psychol.* **2007**, *43*, 1428–1446. [CrossRef]

21. Folmer, E.; Koopmans-van Noorel, A.; Kuiper, W. *Curriculumspiegel 2017*; SLO: Enschede, The Netherlands, 2017.

22. Hakkarainen, A.; Holopainen, L.; Savolainen, H. Mathematical and reading difficulties as predictors of school achievement and transition to secondary education. *Scand. J. Educ. Res.* **2013**, *57*, 488–506. [CrossRef]

23. Baddeley, A. *Working Memory*; Oxford University Press: New York, NY, USA, 1986.

24. Diamond, A. Executive functions. *Annu. Rev. Psychol.* **2013**, *64*, 135–168. [CrossRef]

25. Duckworth, A.L.; Seligman, M.E. Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychol. Sci.* **2005**, *16*, 939–944. [CrossRef] [PubMed]

26. García-Madruga, J.A.; Elosúa, M.R.; Gil, L.; Gómez-Veiga, I.; Vila, J.O.; Orjales, I.; Contreras, A.; Rodríguez, R.; Melero, M.A.; Duque, G. Reading comprehension and working memory's executive processes: An intervention study in primary school students. *Read. Res. Q.* **2013**, *48*, 155–174. [CrossRef]

27. Cragg, L.; Gilmore, C. Skills underlying mathematics: The role of executive function in the development of mathematics proficiency. *Trends Neurosci. Educ.* **2014**, *3*, 63–68. [CrossRef]

28. Blair, C.; McKinnon, R.D. Moderating effects of executive functions and the teacher-child relationship on the development of mathematics ability in kindergarten. *Learn. Instr.* **2016**, *41*, 85–93. [CrossRef]

29. Vandenbroucke, L.; Verschueren, K.; Baeyens, D. The development of executive functioning across the transition to first grade and its predictive value for academic achievement. *Learn. Instr.* **2017**, *49*, 103–112. [CrossRef]

30. Diamond, A. Activities and Programs That Improve Children's Executive Functions. *Curr. Dir. Psychol. Sci.* **2012**, *21*, 335–341. [CrossRef]

31. Resing, W.C.M.; Vogelaar, B.; Elliott, J.G. Children's solving of 'Tower of Hanoi' tasks: Dynamic testing with the help of a robot. *Educ. Psychol.* **2019**, *40*, 1136–1163. [CrossRef]

32. Swanson, H.L. Dynamic testing, working memory, and reading comprehension growth in children with reading disabilities. *J. Learn. Disabil.* **2011**, *44*, 358–371. [CrossRef] [PubMed]

33. Goswami, U.C. Analogical reasoning by young children. In *Encyclopedia of the Sciences of Learning*; Seel, N.M., Ed.; Springer: New York, NY, USA, 2012; pp. 225–228.

34. Richland, L.E.; Simms, N. Analogy, higher order thinking, and education. *WIREs Cogn. Sci.* **2015**, *6*, 177–192. [CrossRef] [PubMed]

35. Peng, P.; Wang, T.; Wang, C.; Lin, X. A meta-analysis on the relation between fluid intelligence and reading/mathematics: Effects of tasks, age, and social economics status. *Psychol. Bull.* **2019**, *145*, 189. [CrossRef]

36. Hessels, M.G.P.; Vanderlinden, K.; Rojas, H. Training effects in dynamic assessment: A pilot study of eye movement as an indicator of problem solving behaviour before and after training. *Educ. Child. Psychol.* **2011**, *28*, 101–113. [CrossRef]

37. Touw, K.W.J.; Bakker, M.; Vogelaar, B.; Resing, W.C.M. Using electronic technology in the dynamic testing of young primary school children: Predicting school achievement. *Educ. Technol. Res. Dev.* **2019**, *67*, 443–465. [CrossRef]

38. Stevenson, C.E.; Bergwerff, C.E.; Heiser, W.J.; Resing, W.C.M. Working memory and dynamic measures of analogical reasoning as predictors of children's math and reading development. *Infant Child. Dev.* **2014**, *23*, 51–66. [CrossRef]

39. Gajda, A.; Karwowski, M.; Beghetto, R.A. Creativity and academic achievement: A meta-analysis. *J. Educ. Psychol.* **2016**, *109*, 259–299. [CrossRef]

40. Mourgues, C.; Tan, M.; Hein, S.; Elliott, J.G.; Grigorenko, E.L. Using creativity to predict future academic performance: An application of Aurora's five subtests for creativity. *Learn. Individ. Differ.* **2016**, *51*, 378–386. [CrossRef]

41. Guilford, J.P. *The Nature of Human Intelligence*; McGraw Hill: New York, NY, USA, 1967.

42. Beghetto, R.A. Creative learning: A fresh look. *J. Cogn. Educ. Psychol.* **2016**, *15*, 6–23. [CrossRef]

43. Vasylkevych, Y.Z. Creativity and intelligence of primary school children: Features of interrelation. *Sci. Educ.* **2014**, *9*, 103.

44. Rindermann, H.; Neubauer, A.C. Processing speed, intelligence, creativity and school performance: Testing of causal hypotheses using structural equation models. *Intelligence* **2004**, *32*, 573–589. [CrossRef]

45. Putwain, D.W.; Kearsley, R.; Symes, W. Do creativity self-beliefs predict literacy achievement and motivation? *Learn. Individ. Differ.* **2012**, *22*, 370–374. [CrossRef]

46. Zbainos, D.; Tziona, A. Investigating primary school children's creative potential through dynamic assessment. *Front. Psychol.* **2019**, *10*, 733. [CrossRef] [PubMed]

47. Dumas, D.G.; Dong, Y.; Leveling, M. The zone of proximal creativity: What dynamic assessment of divergent thinking reveals about students' latent class membership. *Contemp. Educ. Psychol.* **2021**, *67*, 102013. [CrossRef]

48. Onderwijsloket. Available online: https://www.onderwijsloket.com/kennisbank/artikel-archief/hoe-zit-het-nederlandse-onderwijssysteem-in-elkaar/ (accessed on 10 August 2022).

49. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*; Lawrence Erlbaum Associates: Hillsdale, NJ, USA, 1988.

50. Nieuwsbegrip. Available online: https://www.nieuwsbegrip.nl/ (accessed on 24 January 2023).

51. W4Kangoeroe. Available online: https://www.w4kangoeroe.nl/kangoeroe/ (accessed on 24 January 2023).

52. Guilford, J.P. Creativity research: Past, present and future. In *Frontiers of Creativity Research: Beyond the Basic*; Isaksen, S.G., Ed.; Bearly Limited: Buffalo, NY, USA, 1987; pp. 33–65.

53. Vogelaar, B.; Sweijen, S.W.; Resing, W.C.M. Gifted and average-ability children's potential for solving analogy items. *J. Intell.* **2019**, *7*, 19. [CrossRef]

54. Alt, M.; Arizmendi, G.D.; Beal, C.R. The relationship between mathematics and language: Academic implications for children with specific language impairment and English language learners. *Lang. Speech. Hear. Serv. Sch.* **2014**, *45*, 220–233. [CrossRef]

55. Miyake, A.; Friedman, N.P. The nature and organization of individual differences in executive functions: Four general conclusions. *Curr. Dir. Psychol. Sci.* **2012**, *21*, 8–14. [CrossRef]
56. Said-Metwaly, S.; Van den Noortgate, W.; Kyndt, E. Approaches to measuring creativity: A systematic literature review. *CTRA* **2017**, *4*, 238–275. [CrossRef]
57. Stevenson, C.E.; Heiser, W.J.; Resing, W.C.M. Working memory as a moderator of training and transfer of analogical reasoning in children. *Contemp. Educ. Psychol.* **2013**, *38*, 159–169. [CrossRef]
58. Tzuriel, D.; Shamir, A. The effects of mediation in computer assisted dynamic assessment. *J. Comput. Assist. Learn.* **2002**, *18*, 21–32. [CrossRef]
59. Stevenson, C.E.; Touw, K.W.J.; Resing, W.C.M. Computer or paper analogy puzzles: Does assessment mode influence young children's strategy progression? *Educ. Child Psychol.* **2011**, *28*, 67–84. [CrossRef]
60. Chua, Y.P.; Don, Z.M. Effects of computer-based educational achievement test on test performance and test takers' motivation. *Comput. Hum. Behav.* **2013**, *29*, 1889–1895. [CrossRef]
61. Attali, Y.; Arieli-Attali, M. Gamification in assessment: Do points affect test performance? *Comput. Educ.* **2015**, *83*, 57–63. [CrossRef]