The effectiveness of alternative education: a comparison between primary Dalton schools and traditional schools on outcomes of schooling

Sins, P.H.M; van der Zee, S.; Schuitema, J.A.

DOI
10.1080/09243453.2021.1987278

Publication date
2022

Document Version
Final published version

Published in
School Effectiveness and School Improvement

SCHOOL EFFECTIVENESS AND SCHOOL IMPROVEMENT
2022, VOL. 33, NO. 2, 169–197
https://doi.org/10.1080/09243453.2021.1987278

The effectiveness of alternative education: a comparison between primary Dalton schools and traditional schools on outcomes of schooling

Patrick H. M. Sins [a,b], Symen van der Zee[a] and Jaap A. Schuitema [c]

[a]School of Education, Saxion University of Applied Sciences, Deventer, the Netherlands; [b]School of Education, Thomas More University of Applied Sciences, Rotterdam, the Netherlands; [c]Research Institute of Child Development and Education, University of Amsterdam, Amsterdam, the Netherlands

ABSTRACT

With approximately 800 schools, alternative education is a substantial part of primary education in the Netherlands. With nearly 400 schools, Dalton education is the largest form of alternative education in the Netherlands. Given the size and popularity of Dalton education, it is rather remarkable that the effects of these schools have hardly been subject of empirical research. The aim of this study is to compare the cognitive and noncognitive outcomes and citizenship competencies of students attending Dalton schools with those of students from traditional schools. The scores of students in kindergarten, third grade, and sixth grade on language, math (cognitive outcomes), wellbeing, self-efficacy, task motivation (noncognitive outcomes), and citizenship competencies were compared. The results from multilevel analyses show that there are hardly any significant differences between Dalton schools and traditional schools on these measures.

ARTICLE HISTORY
Received 11 March 2021
Accepted 27 September 2021

KEYWORDS
Alternative education; Dalton education; school effectiveness; cognitive and noncognitive outcomes; citizenship competencies

Introduction

At the beginning of the 20th century, several initiatives emerged that aimed at changing the educational system (Howlett, 2013; Kliebard, 2004; Ravitch, 2001). Despite the diversity and the many differences between them, reformers found common ground in their pedagogical discontent with traditional schooling that was claimed to be rigid, authoritarian, and too narrowly focused on students' cognitive development. Reformers conjointly advocated for reconsidering the scope and goals of education through reorganizing educational practices (Imelman & Meijer, 1986; Norris, 2004; Röhrs & Lenhart, 1995). Alternative schools were introduced into the educational landscape, such as Waldorf schools, Jenaplan schools, Freinet schools, Dalton schools, and Montessori schools. The common didactical reform in these schools involved four features: individualizing, activating, contextualizing, and socializing (van der Ploeg, 2013a). The schools tried to find a better balance between the child and the curriculum by letting pupils work

CONTACT Patrick H. M. Sins p.sins@thomasmorehs.nl
This article has been republished with minor changes. These changes do not impact the academic content of the article.

in their own pace and adjusting instruction and assignments to their capabilities and needs (individualizing), by stimulating exploration inside and outside the classroom (activating), by providing opportunities to interact with each other, by stimulating group work (socializing), and by connecting subject matter to pupils' interests and out-of-school experiences (Imelman & Meijer, 1986; Norris, 2004; Röhrs & Lenhart, 1995; van Hulst et al., 1970).

Although the alternative school movement did not immediately result in all-encompassing changes in education, interest in alternative education in the Netherlands remained and even strongly intensified in the 1960s and 1970s (Bakker et al., 2006; van Hulst et al., 1970). By the 1980s, alternative education was steadily gaining in popularity in the Netherlands, and the number of alternative schools kept increasing during the 1990s and after the change of the millennium (Bakker et al., 2006). Currently, there are about 800 alternative schools for primary education in the Netherlands, and numbers continue to increase. The popularity of alternative education is also rapidly growing in other countries over the world (de Bilde, 2013; Lillard, 2019). Given the size and appreciation of alternative education, it is striking that little research has been conducted on its effectiveness. The goal of the present study is to investigate the added value of Dalton primary schools in the Netherlands by comparing the outcomes of their students with those of students attending traditional schools. We will compare the outcomes of pupils in kindergarten, third grade, and sixth grade on language, reading, math (cognitive outcomes), wellbeing, self-efficacy, and task motivation (noncognitive outcomes; see van der Wal & Waslander, 2007) and citizenship competencies.

Effectiveness of alternative education

Alternative schools have a long history in many Western countries but have rarely been subject of effectiveness research. Only few studies have been conducted so far. One of the most comprehensive studies in this respect is that of de Bilde (2013). In four studies, she examined the impact of alternative schools for primary education in Flanders (the Dutch-speaking part of Belgium) in terms of language, math, academic motivation, and engagement. The development of pupils in these areas was analyzed using multilevel growth curve modeling controlling for background characteristics, such as socioeconomic-status, initial achievement, gender, and age. Results indicate that pupils at alternative schools perform significantly and substantially lower than children in regular schools on math, academic motivation, as well as on engagement.

The findings of de Bilde (2013) can be considered alarming for alternative education (see also de Bilde et al., 2013). However, an important limitation of the research conducted by de Bilde et al. is that in all studies the alternative schools were examined as one group. Freinet schools and Waldorf schools were combined to constitute the variable "Alternative Schools". The combined approach involves an oversimplified view of alternative education, as de Bilde (2013) also notes herself. Not only do Freinet and Waldorf schools structurally differ in their underlying principles, moreover, there are substantial differences between these schools in how subject matter is selected and taught to pupils (Imelman & Meijer, 1986). These differences complicate de Bilde's (2013) findings since it remains unclear whether the concerns about the effectiveness of alternative education would apply to all schools equally.

The few studies that have focused on investigating specific forms of alternative education indicate that there are indeed differences in effectiveness. The two alternative educational approaches that have been studied are Montessori education (Dohrmann et al., 2007; Lillard, 2012; Lillard & Else-Quest, 2006; Lillard et al., 2017) and Waldorf education (Cunningham & Carroll, 2011; Steenbergen, 2009). Cunningham and Carroll (2011) compared 30 pupils in first grade (ages 7–9) from two Waldorf schools in the United Kingdom with a matched group of 31 pupils from a traditional school (ages 4–6). Even though children at the traditional school were on average significantly younger than the children at the Waldorf schools, they outperformed them on spelling. No differences were found on reading, phonological awareness, and letter knowledge. Steenbergen (2009) studied the effectiveness of Waldorf secondary schools in the Netherlands by comparing a set of cognitive and noncognitive outcomes. Results show that the scores of pupils in Waldorf schools were significantly lower on mathematics, language, as well as general problem solving. The results on the noncognitive measures were less unequivocal: Students in Waldorf schools scored significantly higher on openness, academic self-image, relationship with teachers, and the use of learning strategies, but they had significantly lower scores on extraversion, conscientiousness, and emotional stability.

The results from the research on Montessori schools are rather different. A study conducted by Lillard and Else-Quest (2006) indicates that pupils at Montessori schools performed better on standardized tests of reading and math, engaged more in positive interaction on the playground, and showed more advanced social cognition and executive functions at the end of kindergarten compared to pupils in traditional schools. Moreover, at the end of elementary school, Montessori children wrote more creative essays, performed better on a task for social dilemma's, and reported feeling a higher sense of community at their school. A more recent study by Lillard et al. (2017) corroborates these findings. Children in kindergarten were tested four times over 3 years from the first semester to the end of preschool (ages 3–6) on measures of academic achievement, executive functioning, social understanding, mastery orientation, and liking of academic tasks. The children in Montessori preschools outperformed the children in the conventional schools on all measures (see also Lillard, 2012). In addition to these findings, Dohrmann and others (2007) report that students in American high schools who had attended Montessori programs from preschool through fifth grade performed significantly better on math and science as compared to a matched group of students who had attended traditional education.

Research on the effectiveness of the other three progressive schools, Jenaplan, Freinet, and Dalton education, is scarce. No studies have specifically examined the effectiveness of Freinet schools, and only one exploratory study has been carried out on the effectiveness of Jenaplan education. Paas and Mulder (2010) compared the school performances of children at Jenaplan schools in kindergarten, third grade, and sixth grade on language and math with those of children in conventional schools. Results show that the outcomes of children in all grades at Jenaplan schools are significantly lower on language and mathematics. Sins and van der Zee (2015) investigated the differences in cognitive and noncognitive outcomes between Dalton schools and traditional schools in primary education (see also van der Zee, 2015). We compared the scores of students in kindergarten, Grade 3, and Grade 6 on language, reading, mathematics (cognitive outcomes), well-

being, cognitive self-esteem, task motivation (noncognitive outcomes), and citizenship competencies, on the basis of data that were collected in a Dutch national cohort study (COOL$^{5\text{-}18}$; see Driessen et al., 2009). Most of our analysis showed no significant differences between the scores of students attending Dalton schools and those from traditional schools. Only one significant difference was found: Students in kindergarten at Dalton schools score significantly higher on language compared to pupils in traditional schools.

Considerations on research on alternative schools

The research on alternative schools is not without its limitations. An important issue concerns the assessment of implementation of the educational concept. Most of the aforementioned studies did not check to what extent schools put their particular educational concept into practice. Instead, these studies assumed that the alternative schools implement the educational concept they claim to carry out and that they do so in a proper manner (Cunningham & Carroll, 2011; de Bilde, 2013; Dohrman et al., 2007; Lillard & Else-Quest, 2006; Paas & Mulder, 2010; Steenbergen, 2009). This assumption is problematic since the actual practice at alternative schools might not correspond with the pedagogical approach the schools claim to adhere to. de Bilde is aware of this conceivable problem and stresses the importance of doing an implementation fidelity check (see also de Bilde et al., 2013).

Such a check has been carried out by Lillard (2012) in a study on the effectiveness of Montessori education (see also Lillard et al., 2017). On the basis of classroom observations, she determined whether the fidelity of implementation of the Montessori program was either high or low. The school performances of preschoolers in the high- and low-fidelity classes were compared with each other and with those of children in a traditional educational setting. Results show that the quality of implementation matters. Children in the high-fidelity group showed significantly larger gains on outcome measures of executive functioning, reading, math, vocabulary, and problem solving, as compared to children in the conventional program and the children in the low-fidelity Montessori classes. The research by Lillard underscores the importance of determining the fidelity of implementation.

In addition, it is not always clear how education is organized in the traditional schools that are used to compare with alternative schools. For instance, Steenbergen (2009) compares Waldorf schools for secondary education with “regular schools”, and Lillard and Else-Quest (2006), Lillard (2012), and Lillard et al. (2017) compare the outcomes of Montessori schools with “non-Montessori” schools. What the educational approach at the “average” or “non-Montessori” school is and how it is implemented in practice at these schools is unclear (cf. Cunningham & Carroll, 2011; Dohrman et al., 2007; Paas & Mulder, 2010). Although it is claimed that the educational practice at regular schools and alternative schools differs, these studies do not offer empirical data to support this claim (see also de Bilde, 2013).

In the present study, we specifically focus on examining the effectiveness of Dutch primary Dalton schools. The quality of implementation of the Dalton approach will be checked, and a comparison will be made with traditional schools that clearly have a different educational concept.

Dalton education in the Netherlands

Dalton education has its roots in the work of the American teacher Helen Parkhurst (1922) and attracted a great deal of international interest, and within a few months the Dalton plan was being implemented in England, Australia, New-Zealand, Germany, India, China, and the Netherlands (van der Ploeg, 2013a, 2013b, 2014). Only in the Netherlands did it continue to exist, even though up until the late 1970s there were only a modest number of schools. From the mid-1980s onwards, Dalton education experienced a growth spurt, and with nearly 400 schools it is now by far the largest reform pedagogical movement in the Netherlands.

Although Dalton education in the Netherlands has its roots in the work of Parkhurst, during the 20th century practitioners continuously kept adapting the plan (van der Ploeg, 2014). Characteristic for present-day Dutch Dalton education are the principles: freedom, independence, collaboration, and the assignments (de Haan, 2015; van der Zee, 2015). Freedom implies that pupils can work at their own pace, plan their work themselves, and choose where they want to work. Independence means that children work on assignments that have been attuned to their learning capabilities and pace of work. The teacher supports autonomous learning by providing instructions and feedback when necessary. The principle collaboration implies interaction and cooperation. Children are allowed to freely interact and work with one another on assignments. Furthermore, children regularly work in groups, and teachers teach the children how to collaborate in a social and efficient manner.

In this study, we examine the effectiveness of Dutch Dalton schools for primary education. We compare the cognitive and noncognitive outcomes and citizenship competencies of pupils at Dalton schools with those of pupils at traditional schools controlling for outcomes obtained from a previous cohort (see Sins & van der Zee, 2015; van der Zee, 2015). School performances of children in kindergarten, third grade, and sixth grade are being compared using hierarchical linear modeling (HLM). The research question is: *Are there differences in effectiveness with respect to cognitive and noncognitive outcomes and citizenship competencies between Dalton schools and traditional schools, after controlling for students' background characteristics and previous achievement scores?*

Method

Data

To determine the main value added of Dalton schools for primary education, data were included from a large-scale triennial cohort study in the Netherlands named COOL$^{5\text{-}18}$. The COOL study follows students aged 5 to 18 throughout their educational careers in primary, secondary, and vocational education with a total sample size of 36,060 students (cf. Hornstra et al., 2015; Karssen et al., 2016). For the present study, we analyzed the data of primary school children from kindergarten, Grade 3, and Grade 6 who were included in the second COOL$^{5\text{-}18}$ measurement. In addition, to control for previous cognitive and noncognitive outcomes of these students, we made use of the data that were obtained from these students during the first wave of COOL$^{5\text{-}18}$. Table 1 shows the number of students in Dalton and in traditional schools included in our subsample as well as their background

Table 1. Overview of the variables, instruments, and number of schools and pupils in the analyses.

	Covariates	Instrument	Dependent variables	Instruments	Number Dalton/traditional schools (number of pupils)
Background charateristics	Gender (boy/girl)				
	Age (in months)				
	Parental educational level (lower secondary vocational; vocational; higher education/ university)				
	Ethnicity (ethnic minority/Dutch background)				
	Parental involvement	Scale parental involvement (PRIMA)			
Kindergarten			Math	Cito categorization	10(113)/5(81)
			Language	Cito language	12(149)/5(80)
Third grade	Intelligence	Noncognitive Capacities Test (van Batenburg & van der Werf, 2004)			
	Math	Cito math (from kindergarten in first COOL$^{5\text{-}18}$ wave)	Math	Cito math	6(59)/8(107)
	Language	Cito language (from kindergarten in first COOL$^{5\text{-}18}$ wave)	Vocabulary	Cito vocabulary	7(75)/8(108)
			Decoding	Cito decoding	7(51)/9(123)
			Reading	Cito reading comprehension	5(55)/9(129)
			Wellbeing teacher	Peetsma et al. (2001)	7(76)/9(129)
			Wellbeing other pupils	Peetsma et al. (2001)	7(75)/9(129)
			Self-efficacy	Midgley et al. (2000)	7(75)/9(129)
			Task motivation	Seegers et al. (2002)	7(75)/9(129)
Sixth grade	Intelligence	Noncognitive Capacities Test (van Batenburg & van der Werf, 2004) (from third grade in first COOL$^{5\text{-}18}$ wave)			
	Language	Cito vocabulary (from third grade in first COOL$^{5\text{-}18}$ wave)	Vocabulary	Cito vocabulary	5(60)/4(55)
		Cito decoding (from third grade in first COOL$^{5\text{-}18}$ wave)	Decoding	Cito decoding	10(115)/8(124)
			Cito test	Cito test	5(67)/7(102)
	Wellbeing in relation to teacher	Wellbeing teacher (from third grade in first COOL$^{5\text{-}18}$ wave)	Wellbeing teacher	Peetsma et al. (2001)	12(132)/10(148)

Wellbeing in relation to students	Wellbeing other pupils (from third grade in first COOL$^{5\text{-}18}$ wave)	Wellbeing other pupils	Peetsma et al. (2001)	12(132)/10(148)
Self-efficacy	Self-efficacy (from third grade in first COOL$^{5\text{-}18}$ wave)	Self-efficacy	Midgley et al. (2000)	12(132)/10(148)
Task motivation	Task motivation (from third grade in first COOL$^{5\text{-}18}$ wave)	Task motivation	Seegers et al. (2002)	12(130)/10(148)
		Citizenship – attitude	ten Dam et al. (2011)	11(132)/8(122)
		Citizenship – skills	ten Dam et al. (2011)	11(132)/8(122)
		Citizenship – reflection	ten Dam et al. (2011)	11(131)/8(122)
		Citizenship – knowledge	ten Dam et al. (2011)	11(131/8(124)

characteristics and their scores on cognitive and noncognitive outcomes during the first COOL$^{5\text{-}18}$ measurement.

Selection of schools

In the COOL$^{5\text{-}18}$ study, schools indicated the educational concept they employ. School could choose from the following options: "Montessori", "Jenaplan", "Dalton", "Waldorf", "Freinet", "experiental", "developmental", "authentic learning", "collaborative learning", and "self-regulated learning". If none of the aforementioned alternative concepts applied, schools could select "other, namely" and were asked to explain which educational concept(s) they implemented. For each of the concepts selected, schools were asked to indicate the extent to which they implemented that concept choosing between "partly" or "fully".

Only schools that indicated to "fully" implement the Dalton concept were selected for this study. To assure that the implementation of the concept at these schools is up to standards, we selected 12 schools that were officially certified by the Dutch Association for Dalton schools as licensed Dalton schools. Teachers as well as school principals in these schools follow extensive professional development programs, and their schools are reviewed on a regular basis by an independent group of experts on Dalton education to assess the extent to which they meet the requirements for being a certified Dalton school (see Berends & Wolthuis, 2014).

The selection of traditional schools was also based on the data schools provided with respect to the educational concept they employ. Schools were labeled as "traditional" when they met two selection criteria. First, schools indicated that they did *not* implement any of the given alternative concepts or elements thereof. Second, the schools themselves consider their education to be traditional. These schools selected the option "other, namely" and provided one of the following descriptions to characterize their educational concept: "whole class teaching", "direct teaching", "traditional education", and/or "old learning". As a result of this strict procedure, 10 schools were selected for the traditional group.

Operationalization of the variables

In the COOL$^{5\text{-}18}$ study, different variables were measured at student level. Table 1 provides an overview of the variables from the COOL$^{5\text{-}18}$ study that were used in this study and the instruments that were employed to collect the data (for psychometric information on the COOL$^{5\text{-}18}$ study, see Driessen et al., 2012).

Background characteristics and previous cognitive and noncognitive outcome scores

Schools and parents provided background information of the students who participated in the COOL$^{5\text{-}18}$ study. The following background characteristics are part of the present study: *age*, *gender*, *intelligence*, *ethnicity*, *parents' educational level*, and the level of *parental involvement*. Information about *age*, *gender*, *ethnicity*, and *parents' educational level* came directly from the school administration. *Parental involvement* was measured with a validated scale that had been developed and tested in a previous Dutch

nationwide cohort study called PRIMA (cf. Peetsma et al., 2006). Teachers indicated the level of parental involvement by answering the following three questions: "in this family the parents are involved in school", "in this family learning and curiosity is stimulated", and "in this family parents support their child's learning". *Intelligence* was only measured for students in third grade, employing the Noncognitive Capacities Test (van Batenburg & van der Werf, 2004). This test is comprised of five scales and consists of 85 items in total. The scales are: figure preparation, exclusion, number series, categories, and analogies.

The longitudinal nature of COOL$^{5\text{-}18}$ made it possible to include students' previous outcome scores from the first wave of COOL$^{5\text{-}18}$, 3 years earlier, in our analyses. This means that previous outcome scores for students in third grade comprised of their results on language and math they obtained in kindergarten during the first wave of COOL$^{5\text{-}18}$. For the children in sixth grade, their outcomes on language and math tests in third grade were included in our analyses of the corresponding cognitive outcome measures. Students' scores on self-efficacy, wellbeing, and motivation from the first wave of the COOL$^{5\text{-}18}$ were included in the analyses of the corresponding noncognitive outcome measures (see Table 1). No previous outcome scores were available for children in kindergarten since they were not included in the first COOL$^{5\text{-}18}$ measurement.

Cognitive outcome measures

Cognitive outcomes were measured with test scores on language, reading comprehension, and mathematics achievement from the Dutch National Institute for Educational Measurement (Cito). In kindergarten, two tests were administered: language and math. The language test measures children's proficiency in emergent literacy. The math test measures three types of skills: (1) classifying, (2) sequencing and comparing, and (3) counting. In third grade, tests on language, reading, and math have been administered. Two tests by the Cito have been used to measure children's language skills. The first test measures children's vocabulary. The second test assesses the ability of children to decode words of varying difficulty within a fixed amount of time. Reading was measured with a reading comprehension test. The math test covered several topics, such as number relations, adding and subtracting, multiplying and dividing, complex applications (executing multiple operations), geometry, and calculating with time and money. In sixth grade, tests for vocabulary and decoding were administered that are comparable to those used in third grade, only adapted to the learning level of the pupils. In addition to these tests, we used the results from the national Cito test which Dutch primary school pupils take in the final year of primary education. In this test, pupils are tested in language, mathematics, world orientation, and study skills. World orientation deals with topics related to history, geography, and biology. Study skills deals with understanding study texts, using information resources, and reading tables, diagrams, graphs, and maps. The score on the Cito test is aggregated into a composite learning achievement score.

Noncognitive outcome measures

Questionnaires for measuring noncognitive outcomes of schooling have been administered in third and sixth grade in both waves of the COOL$^{5\text{-}18}$. The following outcomes were used in this study: *wellbeing*, *self-efficacy*, and *task motivation*. *Wellbeing* was measured with a test developed by Peetsma et al. (2001). The test measures wellbeing

in relation to the teacher (7 items) as well as in relation to other fellow students (6 items). *Self-efficacy* was measured with a scale from the Patterns of Adaptive Learning Survey (PALS; Midgley et al., 2000). The Goal Orientation Questionnaire of Seegers et al. (2002) was used to measure *Task motivation.* In answering the questions on wellbeing, self-efficacy, and task motivation, pupils had five response options, ranging from "absolutely not right" to "totally right". The reliability coefficients of the scales on wellbeing in relation to the teacher, wellbeing in relation to other pupils, self-efficacy, and task motivation are .79, .76, .79, and .74, respectively.

Citizenship competencies

In COOL$^{5\text{-}18}$, the citizenship competencies of pupils in the sixth grade were measured using the Citizenship Competence Questionnaire developed by ten Dam et al. (2011). Four scales are used to assess citizenship competencies: knowledge, reflection, skills, and attitudes. The questionnaire operationalizes citizenship in performing four social tasks: democratic performance, societal responsibility, handling conflicts, and dealing with differences. ten Dam and colleagues coupled the four scales with the four social tasks and developed questions for each social task within each scale. The total test consists of 94 items, of which 27 test knowledge, 28 reflection, 15 skills, and 24 attitude.

Analysis

In this study, we investigate the added value of Dalton education in comparison to traditional schooling with respect to cognitive and noncognitive outcomes of schooling as well as citizenship competencies. Multilevel analyses were performed to analyze the data (Hox, 2010; Opdenakker & Van Damme, 2000; Snijders & Bosker, 2011). This method of analyses considers the nested structure of the data by breaking down the total variance and allocating it across the different levels. Three-level analyses are common in research on school effectiveness (pupil-class-school); however, in our data the class and school level proved to coincide. Consequently, our analyses consist of two levels (pupil/school). Outcomes on the math and reading comprehension tests for pupils in sixth grade could not be analyzed due to insufficient data.

In each of the analyses, we compared three models employing the scores obtained on the measures for determining the cognitive and noncognitive outcomes and citizenship competencies obtained in the second wave of the COOL$^{5\text{-}18}$ as dependent variables. In the first model, pupil characteristics (age, gender, ethnicity, and intelligence) were included in the model. In addition, the corresponding cognitive and noncognitive outcomes from the first wave were added as covariates in this model. In the second model, we added parental characteristics: educational level of the parents and parental involvement. In the final model, the condition, the type of school, was added. If school type is significantly related to student outcomes, this means that there are differences in the mean added value between Dalton schools and traditional schools. We also specified interaction terms to determine the differential added value of the school types for students with different background characteristics. The interaction terms were added to the final model if they contributed significantly to the fit of the model.

We analyzed the data using MLWin (Version 2.30). The estimation method used in all the models is iterative generalized least squares (IGLS). To determine whether adding the

predictors significantly improves the model fit, chi-square-difference tests were performed.

Results

Table 2 provides the descriptive data of our sample's background characteristics and previous outcome scores (scores obtained from the first wave of the COOL$^{5\text{-}18}$ study) of students in kindergarten, Grade 3, and Grade 6 in Dalton schools and in traditional schools.

Chi square analyses reveal no significant differences in the distribution of gender between Dalton schools and regular schools for all grades. In addition, age of students in kindergarten, Grade 3, and Grade 6 did not differ significantly between Dalton schools and traditional schools.

With respect to previous cognitive outcomes, we find that students in Dalton schools score significantly higher than students in traditional schools in third grade (respectively t (204) = 6.02, $p < .01$ for emergent literacy and t (164) = 2.41, $p < .05$ for math. For students in Grade 6, we find no significant differences between Dalton schools and traditional schools with respect to their scores on vocabulary and decoding and on previous-noncognitive outcomes. Also, differences in scores on intelligence in Grade 3 and Grade 6 are not significant.

For students in all three grades, we find that there is a significant difference between Dalton schools and traditional schools with respect to parents' educational level (respectively χ^2 (2, N = 250) = 23.24, $p < .01$ for kindergarten; χ^2 (2, N = 208) = 13.81, $p < .01$ for Grade 3; χ^2 (2, N = 283) = 11.40, $p < .01$ for Grade 6) and ethnicity (respectively χ^2 (1, N = 250) = 66.41, $p < .01$ for kindergarten; χ^2 (1, N = 208) = 25.31, $p < .01$ for Grade 3; χ^2 (1, N = 283) = 27.13, $p < .01$) for Grade 6). This means that parents of students on Dalton schools are significantly more often of Dutch origin and higher educated than partners of students attending traditional schools. In addition, parental involvement in third and sixth grade is significantly higher in Dalton schools compared to traditional schools (respectively t (204) = 3.78, $p < .01$ for Grade 3 and t (281) = 4.90, $p < .05$ for Grade 6). No significant difference was found in parental involvement between Dalton and traditional schools in kindergarten.

Table 3 provides the descriptive results of cognitive outcomes and noncognitive outcomes of students in kindergarten, Grade 3, and Grade 6 in Dalton and traditional schools and citizenship competencies of students in sixth grade.

Cognitive outcomes

Table 4 shows the results of our analyses with respect to students' scores on language and math in kindergarten. For language, adding pupil characteristics results in a significant increase of the model fit. Model fit significantly increases when parental characteristics are added for both outcome scores on language as well as math. The analyses reveal that girls in kindergarten score significantly higher on the language test. In addition, students' age is significantly positively related to their scores on language and math. Also, students score significantly higher on both language and math tests when their parents are of Dutch origin and when parental involvement is high. Finally, students of more highly educated parents score significantly higher both on language and math.

Table 2. Background characteristics and previous outcome scores (scores obtained from the first wave of the COOL[5-18] study) of students in kindergarten, Grade 3, and Grade 6 in Dalton schools and traditional schools.

	Kindergarten			Grade 3			Grade 6		
	Dalton (N = 168)	traditional (N = 82)	p	Dalton (N 77)	traditional (N = 131)	p	Dalton (N = 135)	traditional (N = 148)	p
Gender									
Percentage boys	54.8%	50.0%		55.8%	56.6%		49.6%	47.3%	
Percentage girls	45.2%	50.0%	= .48	44.2%	43.5%	= .93	50.4%	52.7%	= .70
Age in months	71.80 (4.77)	71.83 (4.04)	= .19	109.30 (4.35)	108.62 (4.37)	= .79	147.21 (6.29)	146.16 (5.89)	= .93
Previous cognitive outcomes									
Language				79.24 (9.94)	70.80 (9.56)	< .01			
Math				63.68 (12.22)	58.78 (12.69)	< .05			
Vocabulary							90.18 (9.68)	87.71 (9.58)	= .92
Decoding words							57.26 (18.75)	58.65 (17.56)	= .45
Reading									
Previous Noncognitive outcomes									
Wellbeing in relation to teacher							3.82 (0.52)	3.77 (0.57)	= .29
Wellbeing in relation to other pupils							4.24 (0.56)	4.16 (0.64)	= .07
Self-efficacy							3.85 (0.68)	3.73 (0.65)	= .91
Task motivation							4.23 (0.61)	4.21 (0.60)	= .79
Intelligence[a]				62.99 (11.43)	62.15 (12.27)	= .63	61.68 (11.13)	60.57 (11.69)	= .42
Parents' educational level									
Percentage junior vocational education	7.1%	29.3%		10.4%	32.1%		20.7%	29.1%	
Percentage secondary vocational education	48.8%	30.5%		55.8%	48.1%		39.3%	49.3%	
Percentage higher professional education	44.4%	40.2%	< .01	33.8%	19.8%	< .01	40.0%	21.6%	< .01
Ethnicity									
Percentage ethnic minority	4.2%	46.3%		1.3%	29.8%		7.4%	32.4%	
Percentage Dutch	95.8%	53.7%	< .01	98.7%	70.2%	< .01	92.6%	67.6%	< .01
Parental involvement	3.81 (1.08)	3.80 (0.66)	= .57	4.01 (0.84)	3.56 (0.81)	< .01	3.55 (0.85)	3.53 (0.97)	< .05

Note: Means and standard deviations (within parentheses) or percentages are presented.

[a]Intelligence for Grade 3 involves the scores of students on the Noncognitive Capacities Test in the second wave of the COOL[5-18] study, intelligence for Grade 6 involves the scores of students on the Noncognitive Capacities Test in the first wave of the COOL[5-18] study when they were in Grade 3.

Table 3. Cognitive and noncognitive outcomes scores of students in kindergarten, Grade 3, and Grade 6 and citizenship competencies of students in sixth grade in Dalton schools and traditional schools.

	Kindergarten		Grade 3		Grade 6	
	Dalton	traditional	Dalton	traditional	Dalton	traditional
Cognitive outcomes						
Language	66.89 (11.39)	60.38 (7.99)				
Math	58.23 (11.34)	56.27 (11.88)	72.69 (14.96)	69.16 (17.60)		
Reading			27.89 (14.80)	22.04 (14.38)		
Vocabulary			65.17 (13.40)	59.89 (14.88)	116.92 (12.65)	118.05 (14.55)
Decoding words			73.33 (19.96)	73.76 (14.35)	96.56 (14.25)	98.32 (14.28)
Cito					534.15 (8.51)	534.21 (10.22)
Noncognitive outcomes						
Wellbeing in relation to teacher			3.92 (0.55)	3.83 (0.65)	3.88 (0.53)	3.76 (0.64)
Wellbeing in relation to other pupils			4.11 (0.72)	4.06 (0.66)	4.25 (0.62)	4.20 (0.68)
Self-efficacy			3.90 (0.70)	3.88 (0.62)	3.71 (0.62)	3.78 (0.62)
Task motivation			4.19 (0.68)	4.19 (0.65)	3.97 (0.60)	4.04 (0.61)
Citizenship competencies						
Attitude					3.00 (0.37)	3.02 (0.45)
Skills					3.08 (0.37)	3.07 (0.39)
Reflection					2.22 (0.57)	2.34 (0.59)
Knowledge					0.78 (0.15)	0.72 (0.17)

Note: Means and standard deviations (within parentheses) are presented.

Table 4. Multilevel analyses of cognitive outcomes in kindergarten.

	Model 1				Model 2				Model 3			
	Language		Math		Language		Math		Language		Math	
	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*
Intercept	64.238	1.694	57.551	1.373	55.160	2.404	49.347	1.883	52.858	2.662	49.344	1.958
Gender (reference = boy)	1.930	1.213	-0.297	1.245	2.811**	1.149	1.083	1.187	2.768**	1.146	1.082	1.183
Age	0.360**	0.132	0.391**	0.137	0.411***	0.124	0.449***	0.129	0.406***	0.123	0.449***	0.129
Parents' educational level (reference =	vocational education				5.063***	1.839	0.345	1.768	4.850***	1.838	0.343	1.772
lower secondary vocational education)	Higher education/ university				6.705***	1.914	3.868*	1.865	6.519***	1.910	3.998*	1.864
Ethnicity (reference = ethnic minority)					4.465***	1.994	7.504***	1.578	3.601	2.061	7.495***	1.740
Parental involvement					1.884**	0.759	2.371**	0.792	1.990**	0.758	2.371**	0.791
School type (reference = traditional)									4.604	2.702	2.582	2.070
Variance												
Class	36.208	14.747	22.709	10.937	20.794	9.098	5.611	5.887	17.861	8.066	5.624	5.939
Pupil	78.135	7.605	108.775	9.457	67.943	6.613	97.288	8.547	67.688	6.588	97.281	8.495
% variance explained (vs. previous model)	3.0		2.8	22.4		21.7	3.6		0.0			
Improvement of fit (vs. previous model)	9.760		5.239	37.583		32.296	2.789		0.000			
p value	$p < .01$		$p = .07$		$p < .01$		$p < .01$		$p = .09$		$p = 1$	

*$p < .05$. **$p < .01$. ***$p < .001$.

No differences are found in the cognitive outcome scores between students in Dalton schools and traditional schools. Adding the type of school does not result in a significant increase of the model fit. No significant interaction effects are found.

The results of the multilevel analyses for Grade 3 with respect to vocabulary, decoding words, reading, and math are presented in Table 5. Only the addition of student background characteristics (i.e., previous cognitive outcomes, gender, and intelligence) results in a significant improvement of the model fit for vocabulary, reading, and math. For these three measures, significant positive relations are found with students' previous cognitive outcome score on language in kindergarten and students' intelligence. In addition, older students on average score significantly lower on math than younger students. Finally, boys score significantly higher on math compared to girls. No significant improvement in model fit is observed for decoding. Addition of parental background variables (i.e., parents' educational background, ethnicity, and parental involvement) and of school type does not result in a significant improvement of the model fit for all cognitive outcome measures. In addition, interaction effects are found to be not significant.

Table 6 presents the analyses for cognitive outcomes of students in Grade 6. The model fit significantly improves as a result of adding students' background characteristics (i.e., previous cognitive outcome scores, age, and intelligence) for all three dependent variables. Both the respective previous cognitive outcomes are significantly positively related to students' scores on vocabulary and decoding. There is a significant negative relation between age and students' scores on the composite learning achievement score (i.e., Cito). Also, intelligence of students as measured during the first wave of the COOL$^{5\text{-}18}$ study is significantly positively related to their scores on the Cito test. The model fit does not significantly improve after adding the parental background variables and the variable for school type. Also, no significant interaction effects are found.

Noncognitive outcomes

In Table 7, the results of the multilevel analyses of noncognitive outcome measures in Grade 3 are presented. Adding parental background variables resulted in a significant improvement of the model fit of the second models pertaining to wellbeing in relation to the teacher and for self-efficacy. In both models, there is a significant positive relationship between parental involvement, on the one hand, and students' wellbeing in relation to the teacher and their self-efficacy, on the other hand. Girls score significantly higher than boys on wellbeing in relation to the teacher. Students whose parents are of Dutch origin score significantly lower on self-efficacy compared to students whose parents are from an ethnic minority group. There is no significant difference between school types with respect to students' noncognitive outcomes, and no significant interaction effects are found.

Table 8 presents the results of the analyses with respect to noncognitive outcomes of students in sixth grade. For all noncognitive outcomes, model fit significantly increases following the addition of pupil characteristics. Except for wellbeing in relation to teacher, all the previous noncognitive outcomes as measured during the first wave of COOL$^{5\text{-}18}$ – when students were in third grade – are significantly positively related to wellbeing in relation to other students, self-efficacy, and task motivation, respectively. Intelligence measured during the first wave of COOL$^{5\text{-}18}$ was significantly negatively related to

Table 5. Multilevel analyses of cognitive outcome measures in Grade 3.

	Model 1								Model 2							
	Vocabulary		Decoding		Reading		Math		Vocabulary		Decoding		Reading		Math	
	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*
Intercept	61.063	1.782	73.191	2.120	23.053	1.425	78.803	1.532	55.271	2.861	77.980	3.867	20.665	2.640	72.591	2.806
Previous cognitive outcomes																
Language	0.455***	0.100	0.144	0.145	0.247	0.100			0.397***	0.099	0.145	0.152	0.182	0.103		
Math							0.259*	0.079							0.236*	0.079
Gender (reference = boy)	−1.834	1.590	0.282	2.370	1.233	1.751	−9.075***	1.654	−1.661	1.590	0.183	2.366	1.648	1.747	−9.038***	1.630
Age	−0.015	0.191	−0.344	0.284	−0.137	0.217	−0.490*	0.210	0.034	0.191	−0.313	0.284	−0.124	0.220	−0.392	0.207
Intelligence	0.402***	0.075	−0.009	0.108	0.571***	0.083	0.715***	0.079	0.355***	0.078	−0.031	0.113	0.508***	0.088	0.634***	0.093
Parents' educational level (reference = lower secondary vocational education)																
vocational education									3.477	2.135	−1.601	3.008	0.472	2.244	1.097	2.116
higher education/university									0.505	2.521	0.716	3.625	−1.258	2.702	2.890	2.522
Ethnicity (reference = ethnic minority)									5.299	2.817	−5.253	3.628	3.122	2.565	0.157	2.750
Parental involvement									1.455	1.192	2.066	1.857	1.892	1.373	2.989*	1.317
School type (reference = traditional)																
Variance																
Class	23.216	14.498	25.922	17.567	9.160	7.474	15.118	9.463	11.349	7.586	21.561	15.752	4.838	5.588	9.561	7.131
Pupil	104.222	11.356	225.849	25.268	130.961	14.148	104.118	11.930	104.539	11.370	222.577	24.876	130.180	14.050	101.288	11.575
% variance explained (vs. previous model)	20.0		2.1		21.3		41.0		9.7		3.0		3.6		7.0	
Improvement of fit (vs.previous model)	28.899		1.478		44.281		87.104		8.416		3.822		4.432		7.819	
p value	*p* < .001		*p* = .69		*p* < .001		*p* < .001		*p* = .08		*p* = .43		*p* = .35		*p* = .10	

	Model 3							
	Vocabulary		Decoding		Reading		Math	
	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*
Intercept	55.265	2.879	78.704	3.921	20.264	2.640	72.335	2.831
Previous cognitive outcomes								
Language	0.395***	0.102	0.184	0.155	0.152	0.108		
Math							0.231*	0.080
Gender (reference = boy)	−1.654	1.591	0.088	2.357	1.781	1.749	−9.028***	1.629
Age	0.032	0.191	−0.280	0.285	−0.134	0.219	−0.399*	0.208
Intelligence	0.357***	0.080	−0.056	0.115	0.525***	0.091	0.643***	0.085
Parents' educational level (reference = lower secondary vocational education)								
vocational education	3.454	2.140	−1.479	2.996	0.449	2.245	1.088	2.114
higher education/university	0.472	2.541	1.167	3.632	−1.335	2.701	2.871	2.522
Ethnicity (reference = ethnic minority)	5.126	2.986	−4.263	3.728	2.813	2.587	−0.326	2.914
Parental involvement	1.422	1.194	2.321	1.866	1.730	1.384	2.849*	1.336
School type (reference = traditional)	0.363	2.742	−4.711	4.085	2.034	2.564	1.351	2.713
Variance								
Class	11.586	7.687	22.536	16.093	3.842	5.180	9.976	7.291
Pupil	104.431	11.358	220.389	24.636	130.418	14.068	100.945	11.538
% variance explained (vs. previous model)	0.6		0.5		0.6		0.1	
Improvement of fit (vs. previous model)	0.017		1.321		0.596		0.245	
p value	$p = .62$		$p = .25$		$p = .44$		$p = .62$	

*$p < .05$. **$p < .01$. ***$p < .001$.

Table 6. Multilevel analyses of cognitive outcome measures in Grade 6.

	Model 1						Model 2						Model 3					
	Vocabulary		Decoding		Cito		Vocabulary		Decoding		Cito		Vocabulary		Decoding		Cito	
	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*
Intercept	117.661	1.843	98.221	1.172	534.572	1.126	122.278	3.548	97.572	2.144	533.606	2.030	122.619	3.657	98.202	2.226	532.083	2.068
Previous cognitive outcomes																		
Vocabulary	0.597***	0.135					0.661***	0.141					0.656***	0.141				
Decoding			0.567***	0.036					0.571***	0.036					0.572***	0.036		
Gender (reference = boy)	0.020	2.184	-1.061	1.265	0.108	1.124	-0.408	2.196	1.075	1.271	0.017	1.108	-0.420	2.198	-1.051	1.268	-0.034	1.101
Age	-0.097	0.185	-0.129	0.104	-0.316*	0.100	-0.190	0.195	-0.093	0.107	-0.267*	0.101	-0.181	0.198	-0.079	0.107	-0.260*	0.100
Intelligence	0.222*	0.123	0.032	0.057	0.461***	0.051	0.240*	0.124	0.005	0.060	0.416***	0.053	0.246*	0.124	0.003	0.060	0.415***	0.053
Parents' educational level (reference = lower secondary vocational education)																		
	vocational education						-0.380	2.682	-0.431	1.681	-0.473	1.503	-0.542	2.685	-0.561	1.682	-0.486	1.493
	higher education/ university						-4.226	3.472	0.767	1.951	1.597	1.747	-4.282	3.557	0.965	1.956	1.757	1.738
Ethnicity (reference = ethnic minority)							-3.842	3.265	0.764	1.984	1.088	1.853	-3.680	3.228	1.344	2.060	2.383	2.022
Parental involvement							-0.487	1.276	0.764	0.786	1.129	0.739	-0.523	1.284	0.738	0.784	1.041	0.737
School type (reference = traditional)													-0.553	2.667	-2.111	2.063	-3.163	2.161
Variance																		
Class	9.429	9.448	9.297	5.537	7.047	4.421	2.242	6.861	9.610	5.612	7.576	4.561	2.955	6.248	9.558	5.571	7.874	4.660
Pupil	130.227	17.871	90.407	8.587	50.264	5.595	129.587	17.761	89.476	8.500	48.205	5.369	130.570	17.881	89.091	88.465	47.497	5.291
% variance explained (vs. previous model)	23.2		50.8		32.9		4.2		0.6		3.2		0.3		0.4		0.4	
Improvement of fit (vs. previous model)	32.627		177.143		75.118		3.288		2.038		4.152		0.027		1.044		0.265	
p value	$p < .001$		$p < .001$		$p < .001$		$p = .52$		$p = .73$		$p = .39$		$p = .87$		$p = .31$		$p = .61$	

*$p < .05$. **$p < .01$.***$p < .001$.

Table 7. Multilevel analyses of noncognitive outcome measures in Grade 3.

	Model 1								Model 2							
	Wellbeing teacher		Wellbeing pupils		Self-efficacy		Task motivation		Wellbeing teacher		Wellbeing pupils		Self-efficacy		Task motivation	
	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*
Intercept	3.744	0.107	4.055	0.071	3.889	0.077	4.226	0.093	3.800	0.147	4.105	0.141	3.985	0.136	4.368	0.150
Gender (reference = boy)	0.231***	0.073	0.048	0.093	0.061	0.088	-0.028	0.088	0.257***	0.071	0.076	0.092	0.092	0.087	-0.011	0.088
Age	-0.004	0.009	-0.017	0.011	-0.016	0.011	-0.014	0.011	-0.001	0.009	-0.012	0.011	-0.013	0.010	-0.012	0.011
Intelligence	-0.002	0.003	0.007	0.004	0.003	0.004	0.005	0.004	-0.007	0.004	0.003	0.004	-0.002	0.004	0.003	0.004
Parents' educational level (reference = lower secondary vocational education)																
vocational education									-0.009	0.095	-0.037	0.122	0.160	0.115	0.013	0.117
Higher education/ university									-0.103	0.112	0.090	0.141	0.125	0.133	-0.065	0.137
Ethnicity (reference = ethnic minority)									-0.049	0.132	-0.083	0.137	-0.294*	0.132	-0.181	0.144
Parental involvement									0.235***	0.054	0.163*	0.066	0.176**	0.063	0.127*	0.065
School type (reference = traditional)																
Variance																
Class	0.140	0.058	0.018	0.018	0.036	0.024	0.050	0.029	0.108	0.047	0.012	0.016	0.017	0.016	0.047	0.028
Pupil	0.256	0.026	0.425	0.044	0.375	0.039	0.376	0.039	0.237	0.024	0.413	0.042	0.361	0.037	0.367	0.038
% variance explained (vs. previous model)	2.5		3.7		2.1		0.7		12.9		4.0		8.0		2.8	
Improvement of fit (vs. previous model)	0.314		6.515		3.743		3.585		18.701		7.809		12.566		5.104	
p value	$p = .96$		$p = .09$		$p = .29$		$p = .31$		$p < .001$		$p = .10$		$p < .05$		$p = .28$	

(*Continued*)

Table 7. Continued.

	Model 3							
	Wellbeing teacher		Wellbeing pupils		Self-efficacy		Task motivation	
	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*
Intercept	3.793	0.156	4.105	0.141	3.985	0.136	4.367	0.152
Gender (reference = boy)	0.257***	0.071	0.076	0.092	0.092	0.087	-0.011	0.088
Age	-0.001	0.009	-0.012	0.011	-0.013	0.010	-0.012	0.011
Intelligence	-0.007	0.004	0.003	0.005	-0.002	0.004	0.003	0.004
Parents' educational level (reference = lower secondary vocational education)								
vocational education	-0.009	0.095	-0.037	0.122	0.160	0.115	0.013	0.117
higher education/university	-0.103	0.112	0.090	0.142	0.125	0.134	-0.065	0.137
Ethnicity (reference = ethnic minority)	-0.054	0.136	-0.083	0.143	-0.294*	0.138	-0.183	0.151
Parental involvement	0.235***	0.054	0.163*	0.067	0.176**	0.064	0.127*	0.066
School type (reference = traditional)	0.026	0.193	-0.001	0.122	0.000	0.124	0.007	0.156
Variance								
Class	0.108	0.047	0.012	0.016	0.017	0.016	0.047	0.028
Pupil	0.237	0.024	0.413	0.042	0.361	0.037	0.367	0.038
% variance explained (vs. previous model)	0.0		0.0		0.0		0.0	
Improvement of fit (vs. previous model)	0.0		0.0		0.0		0.0	
p value	$p = 1$		$p = 1$		$p = 1$		$p = 1$	

*$p < .05$. **$p < .01$.***$p < .001$.

Table 8. Multilevel analyses of noncognitive outcome measures in Grade 6.

	Model 1								Model 2							
	Wellbeing teacher		Wellbeing pupils		Self-efficacy		Task motivation		Wellbeing teacher		Wellbeing pupils		Self-efficacy.		Task motivation	
	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*
Intercept	3.764	0.067	4.156	0.067	3.728	0.054	3.932	0.064	3.722	0.133	4.161	0.133	4.026	0.101	4.268	0.118
Previous noncognitive outcomes																
Wellbeing in relation to teacher	0.087	0.065							0.099	0.064						
Wellbeing in relation to students			0.236***	0.061						0.248***	0.060					
Self-efficacy					0.232***	0.051							0.231***	0.050		
Task motivation							0.135*	0.058							0.113	0.058
Gender (reference = boy)	0.066	0.067	0.093	0.075	0.043	0.067	0.111	0.071	0.072	0.067	0.087	0.075	0.013	0.067	0.075	0.071
Age	-0.010	0.006	-0.007	0.006	-0.009	0.006	-0.007	0.006	-0.009	0.006	-0.005	0.006	-0.009	0.006	-0.009	0.006
Intelligence	0.003	0.003	-0.006*	0.003	0.012***	0.003	-0.002	0.003	0.003	0.003	-0.008*	0.003	0.013***	0.003	-0.001	0.003
Parents' educational level (reference = lower secondary vocational education																
vocational education									-0.056	0.088	0.069	0.097	-0.064	0.084	-0.006	0.091
higher education/university									-0.170	0.102	-0.020	0.112	-0.055	0.097	-0.099	0.105
Ethnicity (reference = ethnic minority)									0.133	0.119	-0.037	0.121	-0.294***	0.090	-0.355***	0.107
Parental involvement									0.075	0.044	0.128**	0.048	0.042	0.041	0.055	0.045
School type (reference = traditional)																
Variance																
Class	0.045	0.021	0.036	0.020	0.013	0.011	0.034	0.019	0.057	0.025	0.031	0.018	0.000	0.000	0.009	0.011
Pupil	0.294	0.026	0.368	0.032	0.297	0.026	0.327	0.029	0.284	0.025	0.358	0.031	0.295	0.025	0.328	0.029
% variance explained (vs. previous model)	3.7		5.2		13.6		4.0		-0.6		3.7		4.8		6.6	
Improvement of fit (vs. previous model)	9.686		17.963		42.441		10.132		5.769		8.410		11.796		10.778	
p value	$p < .05$		$p < .01$		$p < .01$		$p < .05$		$p = .22$		$p = .08$		$p < .05$		$p < .05$	

(*Continued*)

Table 8. Continued.

	Model 3							
	Wellbeing teacher		Wellbeing pupils		Self-efficacy		Task motivation	
	Coeff.	*SE*	Coeff.	*SE*	Coeff.	*SE*	Coeff.	*SE*
Intercept	3.678	0.136	4.148	0.134	4.025	0.101	4.267	0.114
Previous noncognitive outcomes								
Wellbeing in relation to teacher	0.095	0.064						
Wellbeing in relation to students			0.245***	0.060				
Self-efficacy					0.230***	0.051		
Task motivation							0.103	0.058
Gender (reference = boy)	0.072	0.067	0.086	0.075	0.012	0.067	0.071	0.071
Age	-0.010	0.006	-0.006	0.006	-0.010	0.006	-0.011	0.006
Intelligence	0.003	0.003	-0.007*	0.003	0.013***	0.003	0.000	0.003
Parents' educational level (reference = lower secondary vocational education)								
vocational education	-0.050	0.088	0.071	0.097	-0.063	0.084	-0.001	0.090
higher education/university	-0.178	0.102	-0.026	0.113	-0.062	0.099	-0.126	0.106
Ethnicity (reference = ethnic minority)	0.094	0.121	-0.058	0.125	-0.306***	0.095	-0.425***	0.109
Parental involvement	0.076	0.044	0.130**	0.048	0.043	0.041	0.055	0.044
School type (reference = traditional)	0.147	0.123	0.062	0.111	0.027	0.071	0.135	0.083
Variance								
Class	0.051	0.023	0.029	0.018	0.000	0.000	0.005	0.009
Pupil	0.285	0.025	0.359	0.032	0.295	0.025	0.328	0.029
% variance explained (vs. previous model)	1.5		0.3		0.0		1.2	
Improvement of fit (vs. previous model)	1.364		0.303		0.150		2.562	
p value	$p = .24$		$p = .58$		$p = .70$		$p = .11$	

*$p < .05$. **$p < .01$. ***$p < .001$.

wellbeing in relation to other students and significantly positively related to students' self-efficacy in sixth grade. Adding parental background variables leads to a significant improvement of the model fit for students' self-efficacy and task motivation. Students with Dutch parents score significantly lower than students with partners from an ethnic minority on these variables. The relation between task motivation measured during the first wave of COOL$^{5\text{-}18}$ and students' task motivation in the second wave is not significant in the second model.

Adding school type does not lead to a significant improvement of the model fit, and no significant interaction effects are found.

Citizenship competencies

Table 9 provides the results of the analyses with respect to sixth-grade students' scores on the four scales of the Citizenship Competence Questionnaire. Adding students' gender, age, and intelligence measured in the first wave of COOL$^{5\text{-}18}$ results in a significant improvement of the model fit for attitude and knowledge. Girls score significantly higher compared to boys on both outcome measures. Students' age is significantly negatively related to the knowledge component of citizenship competencies. Intelligence is significantly positively related to knowledge. Adding the parental background variables results in a significant increase of the model fit for all four components. For attitude, skill, reflection, and knowledge, we find that girls score significantly higher than boys. In addition, students with Dutch parents score significantly lower on attitude, skill, and reflection compared to students with parents from an ethnic minority. Parental involvement is significantly positively related to all four components of citizenship competencies.

Adding school type results in a significant improvement of the model fit for knowledge. On this component, students from Dalton schools score significantly higher than students from traditional schools. No significant interaction effects are found.

Discussion

In this study, we examined the average added value of Dalton education analyzing data from the national Dutch cohort study COOL$^{5\text{-}18}$. By only selecting certified Dalton schools, we ensured that the schools in our sample implement the Dalton concept and implementation is up to standards. We compared the cognitive and noncognitive outcomes and the citizenship competencies of Dalton schools with traditional schools (cf. van der Wal & Waslander, 2007). More specifically, we compared the outcomes of pupils in kindergarten, third grade, and sixth grade on math, language, wellbeing, self-efficacy, task motivation, and citizenship competencies. In the multilevel analyses, we controlled for age, gender, ethnicity, parents' education level, parental involvement, intelligence, and previous achievement scores. Our research question was: *Are there differences in effectiveness with respect to cognitive and noncognitive outcomes and citizenship competencies between Dalton schools and traditional schools, after controlling for students' background characteristics and previous achievement scores?*

Results indicate there are no differences between Dalton schools and traditional schools on all cognitive and noncognitive outcomes. In addition to students' scores on language and math, we did not find significant differences between Dalton schools

Table 9. Multilevel analyses of citizenship competencies in Grade 6.

	Model 1								Model 2							
	Attitude		Skills		Reflection		Knowledge		Attitude		Skill		Reflection		Knowledge	
	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*	Coefficient	SE	Coefficient	*SE*
Intercept	2.897	0.046	3.027	0.047	2.193	0.059	0.714	0.020	3.073	0.078	3.207	0.076	2.486	0.100	0.747	0.034
Gender (reference = boy)	0.233***	0.048	0.106*	0.045	0.173*	0.071	0.064***	0.017	0.208***	0.046	0.089*	0.045	0.130*	0.070	0.061***	0.017
Age	-0.002	0.004	-0.001	0.004	0.002	0.006	-0.004***	0.001	-0.000	0.004	-0.000	0.004	0.003	0.006	-0.004***	0.001
Intelligence	0.002	0.002	0.002	0.002	-0.003	0.003	0.005***	0.001	0.001	0.002	0.001	0.002	-0.003	0.003	0.005***	0.001
Parents' educational level (reference = lower secondary vocational education)																
vocational education									-0.011	0.062	-0.037	0.061	-0.043	0.091	-0.032	0.023
higher education/ university									0.016	0.072	-0.068	0.070	0.050	0.104	-0.024	0.027
Ethnicity (reference = ethnic minority)									-0.214***	0.071	-0.177**	0.070	-0.351***	0.090	-0.014	0.030
Parental involvement									0.116***	0.029	0.065**	0.028	0.098*	0.043	0.036***	0.011
School type (reference = traditional)																
Variance																
Class	0.018	0.009	0.021	0.010	0.016	0.013	0.005	0.002	0.010	0.006	0.010	0.006	0.000	0.000	0.003	0.002
Pupil	0.136	0.012	0.122	0.011	0.308	0.028	0.018	0.002	0.126	0.012	0.121	0.011	0.296	0.026	0.017	0.002
% variance explained (vs. previous model)	9.4		1.4		3.0		17.9		11.7		8.4		8.6		4.3	
Improvement of fit (vs. previous model)	25.172		6.179		6.354		65.638		23.917		10.138		19.846		11.984	
p value	$p < .001$		$p = .10$		$p = .10$		$p < .001$		$p < .001$		$p < .05$		$p < .001$		$p < .05$	

	Model 3							
	Attitude		Skills		Reflection		Knowledge	
	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*
Intercept	3 .068	0.080	3.192	0.076	2.494	0.101	0.721	0.034
Gender (reference = boy)	0.208***	0.046	0.088	0.045	0.131*	0.070	0.060***	0.017
Age	-0.001	0.004	-0.001	0.004	0.004	0.006	-0.004***	0.001
Intelligence	0.001	0.002	0.001	0.002	-0.003	0.003	0.005***	0.001
Parents' educational level (reference = lower secondary vocational education)								
vocational education	-0.010	0.062	-0.035	0.061	-0.046	0.091	-0.031	0.023
higher education/university	0.012	0.072	-0.080	0.070	0.065	0.106	-0.026	0.027
Ethnicity (reference = ethnic minority)	-0.222***	0.074	-0.212**	0.071	-0.326***	0.096	-0.033	0.029
Parental involvement	0.117***	0.029	0.067**	0.028	0.098*	0.043	0.036***	0.011
School type (reference = traditional)	0.023	0.069	0.082	0.065	-0.059	0.075	0.075*	0.033
Variance								
Class	0.010	0.006	0.008	0.006	0.000	0.000	0.003	0.002
Pupil	0.126	0.012	0.121	0.011	0.295	0.026	0.017	0.002
% variance explained (vs. previous model)	0.0		1.5		0.3		9.1	
Improvement of fit (vs. previous model)	0.110		1.483		0.609		4.360	
p value	$p = .74$		$p = .22$		$p = .44$		$p < .05$	

*$p < .05$.**$p < .01$.***$p < .001$.

and traditional schools on wellbeing, self-efficacy, and task motivation. We did, however, find one significant difference between Dalton and traditional schools on the knowledge scale of the Citizenship Competence Questionnaire. Students' attending Dalton schools in sixth grade score significantly higher on this component. No significant differences were found on the other three components (i.e., reflection, skills, and attitude).

This research replicates and underscores the results we reported in a previous study, in which most of our analyses showed no significant differences between the cognitive and noncognitive outcomes of students attending Dalton schools and those from traditional schools (Sins & van der Zee, 2015; van der Zee, 2015). Nevertheless, the results of our research are somewhat in contradiction with other school effectiveness research showing that alternative education is less effective than traditional schooling (cf. de Bilde, 2013). Our results offer an interesting counterpart to this suggestion. Traditional education appears to be equally effective compared to education that is based on the reform pedagogical principles freedom, independence, and collaboration. Our findings suggest that different didactical approaches may result in similar cognitive and noncognitive outcomes. This can as well be interpreted as a favorable result for Dalton education because the concerns about the effectiveness of alternative education do not apply to Dalton education. However, to gain more insight into the added value of Dalton schools and into the effectiveness of the alternative principles on which this school type is based, more research is needed, taking into account three limitations of the present study.

To be able to generalize our findings, a larger sample of Dalton and traditional schools needs to be taken. A limitation of the present study was that during the first wave of COOL$^{5\text{-}18}$, about 6% of the total amount of Dalton schools in the Netherlands participated. Between 25% and 62% of the Dalton schools that also participated in the first wave were included in the second wave of COOL$^{5\text{-}18}$. Due to the nature of our sample, the results cannot be generalized to all Dalton schools and traditional schools.

A second limitation concerns the implementation of the Dalton concept in the schools. Earlier, we pointed to the importance of an "implementation fidelity check" (de Bilde et al., 2013; Lillard, 2012). This study addresses this shortcoming in most previous research by only selecting schools that have been certified by the Dutch Association for Dalton schools as licensed Dalton schools and for which the school management has indicated that they have fully implemented the Dalton concept. However, more research is necessary to gain a complete and in-depth understanding of how and to what extent the Dalton principles of freedom, independence, and collaboration have been implemented in the schools.

The final limitation is that this study investigated the average value added of Dalton schools only. On the basis of the present sample, we could not check for differential effectiveness: that is, influences of school and classroom processes that may possibly explain differences in effectiveness. For instance, we found that students attending Dalton schools in sixth grade score significantly higher on the knowledge component of citizenship competencies compared to students in traditional schools. A possible explanation for this finding may be that Dalton schools in upper elementary grades spend more time on stimulating citizenship knowledge, which involves societal knowledge and interpersonal knowledge (Karssen et al., 2016). Future research should provide more insight into these processes in Dalton education.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The authors reported there is no funding associated with the work featured in this article.

Notes on contributors

Patrick H. M. Sins, PhD, is professor of Alternative Education at the Schools of Education at Saxion and Thomas More University of Applied Sciences in the Netherlands. In his research he aims at developing and optimizing innovative perspectives on education based on critical scientific syntheses, on the one hand, and on conducting practice-based scientific research, on the other. He coordinates the research that takes place within his research group focusing on: educational effectiveness, teacher professional development, self-regulated learning, and citizenship education.

Symen van der Zee, PhD, is a professor in Science and Technology Education at the Saxion University. His focus is on the professional development of primary and secondary school teachers with regards to innovative STEM education and the development and testing of educational practices aimed at stimulating and motivating pupils for a career in the STEM fields.

Jaap A. Schuitema, PhD, is an associate professor in the Research Institute of Child Development and Education at the University of Amsterdam. His research interests include student motivation, self-regulated learning, and higher order thinking and reasoning skills. Other areas of research concern the moral development of students and citizenship education. He has studied the effectiveness of teaching strategies and learning environment on students' critical thinking, science skills, moral reasoning, and self-regulated learning and has published several articles in these areas in leading international journals.

ORCID

Patrick H. M. Sins http://orcid.org/0000-0003-0443-6074
Jaap Schuitema http://orcid.org/0000-0002-0125-5537

References

Bakker, N., Noordman, J., & Rietveld-van Wingerden, M. (2006). *Vijf eeuwen opvoeden in Nederland. Idee en praktijk: 1500-2000* [Five centuries of pedagogy in the Netherlands. Idea and practice: 1500-2000]. Van Gorcum.

Berends, R., & Wolthuis, H. (2014). *Focus op dalton* [Focus on Dalton]. Saxion Dalton University Press.

Cunningham, A. J., & Carroll, J. M. (2011). The development of early literacy in Steiner- and standard-educated children. *British Journal of Educational Psychology*, *81*(3), 475–490. https://doi.org/10.1348/000709910X522474

de Bilde, J. (2013). *Alternative education. Examining the effects of alternative educational approaches on student achievement, academic motivation and engagement in Flemish Primary schools* [Doctoral dissertation, KU University Leuven]. Lirias. https://limo.libis.be/primo-explore/fulldisplay?docid=LIRIAS1940949&context=L&vid=Lirias&search_scope=Lirias&tab=default_tab&lang=en_US&fromSitemap=1

de Bilde, J., Van Damme, J., Lamote, C., & De Fraine, B. (2013). Can alternative education increase children's early school engagement? A longitudinal study from kindergarten to third grade. *School Effectiveness and School Improvement*, *24*(2), 212–233. https://doi.org/10.1080/09243453.2012.749795

de Haan, D. (2015). *Schoolverbetering in het daltononderwijs: Onderzoek naar het verband tussen daltononderwijs, schoolverbetering en onderwijskwaliteit* [School improvement in Dalton education: Research on the relationship between Dalton education, school improvement and educational quality] [Unpublished doctoral dissertation]. Rijksuniversiteit Groningen.

Dohrmann, K. R., Nishida, T. K., Gartner, A., Lipsky, D. K., & Grimm, K. J. (2007). High school outcomes for students in a public Montessori program. *Journal of Research in Childhood Education*, *22*(2), 205–217. https://doi.org/10.1080/02568540709594622

Driessen, G., Mulder, L., Ledoux, G., Roeleveld, J., & Van der Veen, I. (2009). *Cohortonderzoek COOL 5–18. Technisch rapport basisonderwijs, eerste meeting 2007/08* [Cohort study COOL 5–18: Technical report primary education, first wave 2007/08]. ITS/Kohnstamm Institute.

Driessen, G., Mulder, L., Ledoux, G., Roeleveld, J., & van der Veen, I. (2012). *Cohortonderzoek COOL$^{5\text{-}18}$. Technisch rapport basisonderwijs, tweede meting 2009/10* [Cohort study COOL$^{5\text{-}18}$. Technical report primary education, second wave 2009/10]. SCO-Kohnstamm Institute.

Hornstra, L., van der Veen, I., Peetsma, T., & Volman, M. (2015). Does classroom composition make a difference: Effects on developments in motivation, sense of classroom belonging, and achievement in upper primary school. *School Effectiveness and School Improvement*, *26*(2), 125–152. https://doi.org/10.1080/09243453.2014.887024

Howlett, J. (2013). *Progressive education: A critical introduction*. Bloomsbury Publishing.

Hox, J. (2010). *Multilevel analysis: Techniques and applications*. Routledge.

Imelman, J. D., & Meijer, W. A. J. (1986). *De nieuwe school, gisteren en vandaag* [The new school, yesterday and tomorrow]. Elsevier.

Karssen, M., van der Veen, I., & Volman, M. (2016). Ethnic diversity in schools and bi-ethnic Dutch students' educational outcomes and social functioning. *School Effectiveness and School Improvement*, *27*(4), 613–628. https://doi.org/10.1080/09243453.2016.1189436

Kliebard, H. M. (2004). *The struggle for the American curriculum: 1893-1958* (3rd ed.). Routledge Falmer.

Lillard, A. S. (2012). Preschool children's development in classic Montessori, supplemented Montessori, and conventional programs. *Journal of School Psychology*, *50*(3), 379–401. https://doi.org/10.1016/j.jsp.2012.01.001

Lillard, A. S. (2019). Shunned and admired: Montessori, self-determination, and a case for radical school reform. *Educational Psychology Review*, *31*(4), 939–965. https://doi.org/10.1007/s10648-019-09483-3

Lillard, A. S., & Else-Quest, N. (2006, September 29). Evaluating Montessori education. *Science*, *313* (5795), 1893–1894. https://doi.org/10.1126/science.1132362

Lillard, A. S., Heise, M. J., Richey, E. M., Tong, X., Hart, A., & Bray, P. M. (2017). Montessori preschool elevates and equalizes child outcomes: A longitudinal study. *Frontiers in Psychology*, *8*, Article 1783 . https://doi.org/10.3389/fpsyg.2017.01783

Midgley, C., Maehr, M. L., Hruda, L. Z., Anderman, E., Anderman, L., Freeman, K. E., Gheen, M., Kaplan, A., Kumar, R., Middleton, M. J., Nelson, J., Roeser, R., & Urdan, T. (2000). *Manual for the Patterns of Adaptive Learning Scales (PALS)*. The University of Michigan.

Norris, N. D. (2004). *Promise and failure of progressive education*. Rowman & Littlefield.

Opdenakker, M.-C., & Van Damme, J. (2000). The importance of identifying levels in multilevel analysis: An illustration of the effects of ignoring the top or intermediate levels in school effectiveness research. *School Effectiveness and School Improvement*, *11*(1), 103–130. https://doi.org/10.1076/0924-3453(200003)11:1;1-#;FT103

Paas, T., & Mulder, L. (2010). Dalton doet het beter dan doorsnee school [Dalton performs better than the average school]. *Didaktief*, *6*, 2–4.

Parkhurst, H. (1922). *Education on the Dalton plan*. E. P. Dutton and Company.

Peetsma, T., van der Veen, I., Koopman, P., & van Schooten, E. (2006). Class composition influences on pupils' cognitive development. *School Effectiveness and School Improvement*, *17*(3), 275–302. https://doi.org/10.1080/13803610500480114

Peetsma, T., Wagenaar, E., & de Kat, E. (2001). School motivation, future time perspective and well-being of high school students in segregated and integrated schools in the Netherlands and the

role of ethnic self-description. In J. K. Koppen, I. Lunt, & C. Wulf (Eds.), *Education in Europe: Cultures, values, institutions in transition* (pp. 54–74). Waxmann.

Ravitch, D. (2001). *Left back: A century of battles over school reform*. Touchstone.

Röhrs, H., & Lenhart, V. (Eds.). (1995). *Progressive education across the continents: A handbook*. Peter Lang.

Seegers, G., van Putten, C. M., & de Brabander, C. J. (2002). Goal orientation, perceived task outcome and task demands in mathematics tasks: Effects on students' attitude in actual task settings. *British Journal of Educational Psychology*, *72*(3), 365–384. https://doi.org/10.1348/000709902320634366

Sins, P. H. M., & van der Zee, S. (2015). De toegevoegde waarde van traditioneel vernieuwingsonderwijs: Een studie naar de verschillen in cognitieve en niet-cognitieve opbrengsten tussen daltonscholen en traditionele scholen voor primair onderwijs [The effectiveness of alternative education: A study on the differences in cognitive and noncognitive outcomes between Dalton schools and traditional schools in primary education]. *Pedagogische Studiën*, *92*(4), 254–274.

Snijders, T. A. B., & Bosker, R. J. (2011). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). Sage Publications.

Steenbergen, H. (2009). *Vrije en reguliere scholen vergeleken: Een onderzoek naar de effectiviteit van vrije scholen en reguliere scholen voor voortgezet onderwijs* [Waldorf and traditional schools compared: A study on the effectiveness of Waldorf and traditional schools for secondary education] [Unpublished doctoral dissertation]. The University of Groningen.

ten Dam, G., Geijsel, F., Reumerman, R., & Ledoux, G. (2011). Measuring young people's citizenship competencies. *European Journal of Education*, *46*(3), 354–372. https://doi.org/10.1111/j.1465-3435.2011.01485.x

van Batenburg, T., & van der Werf, M. (2004). *NSCCT: Verantwoording, normering en handleiding* [NSCCT: Validation, norming, and guideline]. GION.

van der Ploeg, P. A. (2013a). *Dalton plan: Origins and theory of Dalton education*. Saxion Dalton University Press.

van der Ploeg, P. A. (2013b). The Dalton Plan: Recycling in the guise of innovation. *Paedagogica Historica*, *49*(3), 314–329. https://doi.org/10.1080/00309230.2012.725840

van der Ploeg, P. A. (2014). The salient history of Dalton education in the Netherlands. *History of Education*, *43*(3), 368–386. https://doi.org/10.1080/0046760X.2014.887792

van der Wal, M., & Waslander, S. (2007). Traditional and non-traditional educational outcomes: Trade-off or complementarity? *School Effectiveness and School Improvement*, *18*(4), 409–428. https://doi.org/10.1080/09243450701712502

van der Zee, S. (2015). *De effectiviteit van daltononderwijs* [The effectiveness of Dalton education] [Unpublished doctoral dissertation]. Rijksuniversiteit Groningen.

van Hulst, J. W., van der Velde, I., & Verhaak, Th. M. (1970). *Vernieuwingsstreven binnen het Nederlandse onderwijs in de periode van 1900-1940* [Educational reform in The Netherlands in the period 1900-1940]. Wolters-Noordhoff.