The effect of systematic academic instruction on behavioural and academic outcomes of students with EBD

van der Kamp, Lidy; Pijl, Sip Jan; Post, Wendy J.; Bijstra, Jan O.; van den Bosch, Els J.

Published in:
Educational Studies

DOI:
10.1080/03055698.2016.1148584

Publication date:
2016

EDUCATIONAL STUDIES, 2016
VOL. 42, NO. 1, 72–84
http://dx.doi.org/10.1080/03055698.2016.1148584

Routledge
Taylor & Francis Group

The effect of systematic academic instruction on behavioural and academic outcomes of students with EBD

Lidy van der Worp-van der Kamp[a], Sip Jan Pijl[b,c], Wendy J. Post[c], Jan O. Bijstra[a] and Els J. van den Bosch[c]

[a]Regional Expertise Centre for Northern Netherlands Cluster 4 – Behaviour, Groningen, The Netherlands; [b]Pedagogical Institute, Norwegian University of Science and Technology, Trondheim, Norway; [c]Faculty of Behavioural and Social Sciences, University of Groningen, Groningen, The Netherlands

ABSTRACT

This study aims to assess the impact of systematic academic instruction on academic progress and behavioural problems of students with emotional and/or behavioural disorders (EBD) in special education. Earlier studies have noted the importance of a systematic approach as well as the significance of focusing on academic instruction instead of on behaviour. On the basis of these studies, it was hypothesised that the amount of teachers' systematic academic instruction positively influences the academic progress of students with EBD as well as their behaviour. The amount of systematic academic instruction by teachers in daily teaching ($N = 88$) and the behaviour and academic progress of a sample of their students ($N = 234$) were measured and analysed by means of multilevel analyses. The results show a significant relationship between systematic academic instruction and academic outcomes. However, academic outcomes taking into account previous academic performance and behavioural outcomes could not be related to systematic academic instruction. Implications for research on the daily practice of special education are discussed.

ARTICLE HISTORY
Received 4 May 2015
Accepted 24 December 2015

KEYWORDS
Instruction; behaviour; systematic; academic; EBD

Introduction

Students with emotional and/or behavioural disorders (EBD) are a serious challenge to their teachers (Kauffman and Landrum 2013). Numerous students in primary and secondary education are considered to have special educational needs (SEN) because of their problematic behaviour and their number has been growing for years. In 2012/2013, 1.4% of all primary students in the Netherlands were classified as having SEN owing to their serious behavioural or psychiatric problems and 60% of these attend special schools for students with EBD (Koopman and Ledoux 2013). These students' problems are seldom limited to behaviour, they often experience academic problems and poor scholastic achievement as well (Nelson et al. 2004; Bos and Vaughn 2006; Vannest, Temple-Harvey, and Mason 2009; Ledoux et al. 2012). In fact, students in special schools for students with EBD have the poorest educational

CONTACT Lidy van der Worp-van der Kamp l.vdworp@renn4.nl

outcomes of any disability group within special education (Bradley, Doolittle, and Bartolotta 2008; Sacks and Kern 2008, 113; Hagaman 2012; Ledoux et al. 2012) and this deteriorates over time (Hagaman 2012).

Although increasing research on the academic development of students with EBD shows that these students seem to be capable of acquiring fair academic skills (Mooney et al. 2005), education seems to fall short in meeting these students' specific needs. The literature provides us with two major causes of this deficiency. Firstly, numerous researchers accentuate the fact that teachers of students with EBD often tend to focus on students' behaviour rather than on their academic skills (Mooney et al. 2003; Reid et al. 2004; Lane et al. 2005; Pianta and Hamre 2009). Confronted with problematic behaviour, teachers frequently tend to switch to behaviour control or disciplinary practices (e.g. time-out, suspension), or even to exclude students from academic instruction (Levy and Vaughn 2002). Secondly, partly due to the fact that teachers pay so much attention to managing disruptive behaviour, the questions regarding what and how students with EBD should be taught are often not carefully considered (Levy and Chard 2001; Nelson, Benner, and Mooney 2008). However, since problem behaviour frequently occurs when students are confronted with non-appropriate tasks, it is very important that the tasks presented do match the special needs of these students (Umbreit et al. 2007). This so-called "tailored education" requires meticulously, systematically designed instruction (Coleman and Vaughn 2000). Unfortunately, teachers often do not plan their instruction systematically. The instructional content is often prescribed by the method and not sufficiently adapted to students' special needs (Wilen, Hutchinson, and Ishler 2008).

Given these omissions, the academic progress of students with EBD could be improved when teachers increase their focus on (1) academic instruction in (2) a systematic manner. Both dimensions are essential in teaching academic skills to students with EBD (Van der Worp-van der Kamp et al. 2015). Moreover, as a consequence of the above-mentioned relation between academic learning and behaviour, systematic academic instruction may also result in decreases in challenging and disruptive behaviour (Lee, Sugai, and Horner 1999). Yell, Busch, and Rogers (2014) even refer to quality academic instruction as probably the most desirable and economical prevention and intervention strategy for EBD. Although an increasing number of studies seems to confirm this (van der Worp-van der Kamp et al. 2014), caution is necessary. Although much research on academic interventions has reported positive outcomes for students with EBD, it is important to be aware that the majority of these studies were conducted with a relatively small number of students. Moreover, the interventions examined were conducted under controlled circumstances ensuring that the interventions were implemented as intended. Researchers and their assistants were often closely involved in the realisation of the intervention. Generalising of these findings requires research with large sample sizes, examining the focus and feasibility of teachers to teach academic skills in their own classrooms (Therrien et al. 2014).

Since studies on academic instruction to students with EBD are often limited to a few participants in controlled settings, it is important to shift the focus of research to daily educational settings without researchers being involved. The main purpose of this study is to assess the extent to which systematic as well as academic instruction affects students' academic progress as well as their behaviour in daily practice in special education. Based on the literature (vide supra), it is assumed that teachers with high scores for both dimensions generate higher academic and behavioural outcomes compared to teachers with low scores for one or on both dimensions.

Method

Design

A cross-sectional observational study was designed to ascertain the relationship between systematic and academic teaching (independent variables) and behaviour and academic performance of students with EBD (dependent variables). The study took place in six special primary schools for students with severe behavioural problems in the Northern Netherlands (RENN4). All the teachers at these six schools ($N = 88$) were approached to participate in this study. The amount of systematic academic instruction was assessed by two self-report questionnaires. The teachers were also asked to complete a questionnaire assessing their students' behaviour. To limit the burden on teachers, the selection design ensured that teachers only had to provide information on the behaviour of up to five of their students ($N = 328$), 33% of the total population of the six schools. These students were selected on the basis of their last names: the first five students alphabetically were selected per group. In the case of duo teachers, the questionnaire was completed by the main teacher. Finally, information about the student's academic development was obtained from existing data-sets. Paragraph: use this for the first paragraph in a section, or to continue after an extract.

Participants

Sixty-nine teachers (78%) from all of the six schools approached participated in the research. They had on average 13.97 (SD 9.88) years of teaching experience, a substantial part of which in special education ($M = 11.16$, SD = 8.61). The questionnaires returned provided information on 75% ($N = 247$) of the selected students. Since no academic performance can be expected from students <6 years old, these students (5%) were removed from the selection, reducing the number of participating students to 234, 81% of them boys. All the students had been admitted to these schools because they met the following school-specific criteria: (1) Students show severe behavioural or psychiatric problems in terms of DSM-IV; (2) This behaviour is manifested in education as well as in the home and/or leisure activities; (3) Youth care and/or a child psychiatric service were involved with these students; (4) Students' participation in education is extremely limited in terms of serious shortcomings in academic learning and/or behaviour in relation to the teacher or other students; (5) Additional, evident educational care by the school for at least six months generated insufficient progress (WEC-raad 2008). Of the participants, 19.4% were in grade 3/4, 33.3% in grade 5/6 and 47.3% grade 7/8. Their mean age was 10.2 (1.9) and their IQ was 90.7 (15.9) (see Table 1).

Instruments

To measure the degree of systematic academic instruction, Van der Worp-van der Kamp et al. (2015) developed a tool comprising two questionnaires, developed iteratively through a process of empirical input and extensive feedback from teachers. One questionnaire, consisting of 36 items using a four-point Likert response scale, concerned systematic teaching according to the Plan Do Check Act cycle (PDCA). The mean score determined teachers' degree of systematic instruction. Examples of the items included: "I use evaluation data for the preparation of my lessons" and "I adjust the learning goals to the students" capabilities in advance'. This instrument had an α of 0.89. In order to validate the questionnaire, an observation scale

Table 1. Summary of mean scores on PDCA, AI (teacher), SDQ and $CITO_{Maths/Spelling2013}$ and $CITO_{Maths/Spelling2014}$ (student).

	N	Minimum	Maximum	Mean	SD
Teacher					
$PDCA_Q$	55	2.11	3.61	2.85	0.32
AI_Q	55	2.17	3.75	2.67	0.33
Student					
SDQ	234	3	31	15.40	6.46
$CITO_{Maths2013}$	157	1	123	67.50	25.28
$CITO_{Maths2014}$	200	7	145	69.73	27.92
$CITO_{Spelling2013}$	166	83	155	122.54	13.36
$CITO_{Spelling2014}$	197	79	150	124.26	13.36
IQ	183	59	144	90.7	15.9
Age	224	6.1	13.9	10.2	1.9

Notes: $PDCA_Q$ = Questionnaire concerning Plan Do Check Act. AI_Q = Questionnaire concerning academic instruction. $CITO_{Maths/Spelling2013}$/$CITO_{Maths/Spelling2014}$ = Standard biannual tests for math and spelling in June 2013 and January 2014. SDQ = Strengths and Difficulties Questionnaire.

concerning the PDCA was developed. The observation scale included specific descriptions of teacher behaviour concerning the PDCA cycle, also rating from 1 to 4 points. Again, the mean score determined teachers' degree of observed systematic instruction. The validity of the questionnaire was supported by a fair correlation between the outcomes of the questionnaire and the observation scale (0.32). Since a correlation does not automatically imply an agreement between two measurements (Bland and Altman 1986), the validity of the questionnaire was also determined by using a Bland–Altman plot, a graphical techniques to provide information about the agreement and nature of the differences between the questionnaire and the observation scale. Basis on the outcome, the questionnaires were supposed to reflect teachers' systematic academic teaching in a satisfactory manner (Van der Worp-van der Kamp et al. 2015).

The other questionnaire, consisting of 12 items using a four-point Likert response scale, concerned academic instruction vs. behavioural instruction. Examples of the items included "In case of disruptive behaviour, I check the appropriateness of the learning task" and "I mainly reward behaviour". This questionnaire had an α of 0.76. For validation, an observation scale was developed concerning the amount of time spent on academic vs. behavioural instruction. The more teachers spent on behavioural instruction, the lower the score. The outcomes of the correlation between the questionnaire and observation (0.31) as well as the outcomes of the Bland–Altman plot guaranteed the validity of the questionnaire in a satisfying manner (Van der Worp-van der Kamp et al. 2015).

Students' behaviour was measured by the Strengths and Difficulties Questionnaire (SDQ). The SDQ is a brief behavioural screening questionnaire for 3–16 year olds and a psychometrically valid tool for identifying students with various facets of behavioural problems over the last six months (ref). According to Becker et al. (2004), the SDQ total problem score correlates strongly (0.76–0.84) with the more extensive Teacher's Report Form. The SDQ was selected because it is a short questionnaire and teachers had to complete them for up to five students. The instrument involves 25 questions and takes 5–7 min per student to complete. It provides a total scale score (Total Difficulties), and five narrowband subscale scores (Emotional Symptoms, Conduct Problems, Hyperactivity/Inattention, Peer Relationship Problems and Prosocial Behaviour). The Total Difficulties score, based on the addition of

the first four subscale scores, was used for measuring the behavioural development of the students. The Dutch version of the SDQ used has a Cronbach alpha of 0.88 (van Widenfelt et al. 2003). The reliability of the scale was good: 0.79 (Total Difficulties).

The academic performance of the students was measured by the Dutch CITO (Central Institute for Test Development) monitoring and evaluation system, a total assessment programme for school-aged students (2008). This system consists of standard biannual tests for reading, spelling and maths. For this study, the tests for maths ($CITO_{Maths}$) and spelling ($CITO_{Spelling}$) were used. These tests provide an indication of students' performance level by converting raw scores for the tests into ability scores, comparable to a norm group of peers. The maths and spelling ability scores from the end of the preceding academic year ($CITO_{2013}$) were used as controlling variable, and the ability scores halfway through the current academic year ($CITO_{2014}$) were used as dependent variable. The $CITO_{Maths}$ has a Cronbach's alpha of 0.86 or higher, the $CITO_{Spelling}$ a Cronbach's alpha of .90 or higher (CITO 2009).

Procedure

Data collection took place in three successive phases. In the first phase, each teacher was asked to complete the $PDCA_Q$ and the AI_Q. Next each teacher was asked to complete the SDQ for the selected students. All questionnaires were distributed and collected using Qualtrics, an online data collection and analysis system. In the third phase, information about every student's academic development was collected from CITO's LOVS data base.

Data analyses

The descriptive statistics concerning the $PDCA_Q$, AI_Q, SDQ, $CITO_{Maths}$ and $CITO_{Spelling}$ were calculated with SPSS and graphically depicted in box plots. To answer the research question, three multilevel models were computed. One multilevel regression analysis was performed with behaviour (SDQ) as dependent variable and two with academic performance in maths ($CITO_{Maths}$) and spelling ($CITO_{Spelling}$) as dependent variables. In these analyses, teachers were considered as level 2 and students as level 1 to take into account the correlation between students per teacher. The multilevel analysis started with the construction of the most basic multilevel models. These so-called empty models have no predictors, other than the intercept. The models help to find the percentage variance at the different levels without independent variables. In other words, to what extent is the variance in SDQ and CITO scores due to students having different teachers and to what extent to differences between individual students? By then introducing the explanatory variables, the models reveal to what extent the variance at each level is explained by these variables.

Subsequently, models with teachers' systematic academic instruction (AI_Q and $PDCA_Q$) as explanatory independent variables were computed for behaviour (SDQ) and academic performance ($CITO_{Maths2014}$ and $CITO_{Spelling2014}$) as dependent variables. Possible explanatory variables like Gender, Age and IQ were included in the three models. The students' baseline scores ($CITO_{Maths2013}$ and $CITO_{Spelling2013}$) were also included in the models for academic performance. Fixed and random effects were considered. A p-value < 0.05 was considered to be significant. MLwiN 2.23 (Rasbash et al. 2005), a programme specifically designed to analyse hierarchical data, was used to perform the analyses.

Results

Descriptive statistics

Table 1 presents the descriptive scores.

The mean scores per teacher for their systematic academic instruction revealed a small distribution in mean score for the $PDCA_Q$ and AI_Q (Table 1). Box plots (Figure 1) show that the scores on AI are somewhat skewed to the right.

The mean scores were then calculated for all students (Table 1). Concerning the behavioural outcomes, the students' mean score on the SDQ was 15.4 (SD 6.46). The box plot shows a balanced distribution over the students (Figure 2).

Several data were missing from the schools' biannual monitoring and evaluation system (CITO). Percentages of valid scores were 67% ($CITO_{Maths2013}$), 85% ($CITO_{Maths2014}$), 71% ($CITO_{Spelling2013}$) and 84% ($CITO_{Spelling2014}$). Both scores for $CITO_{Maths}$ (63%) and $CITO_{Spelling}$ (68%), necessary to calculate the progression in score, were known for an even smaller number of students. Students' progression in score over six months for $CITO_{Maths}$ was 5.77 (min = –42,

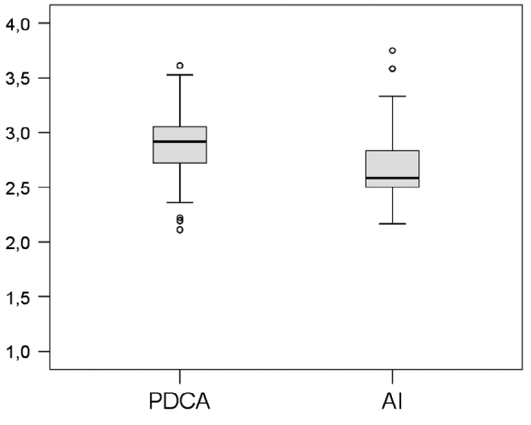

Figure 1. Boxplots concerning teachers' score on the $PDCA_Q$ and AI_Q.

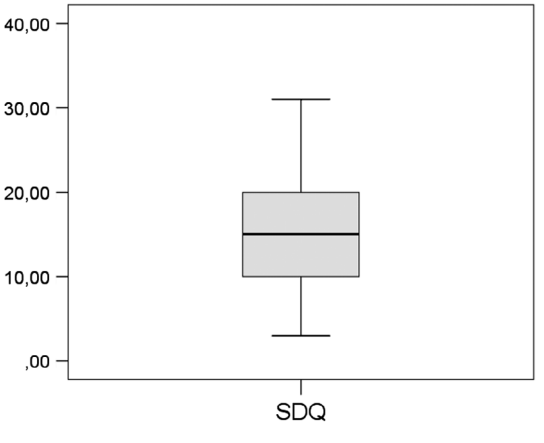

Figure 2. Boxplot concerning students' score on the Strengths and Difficulties Questionnaire.

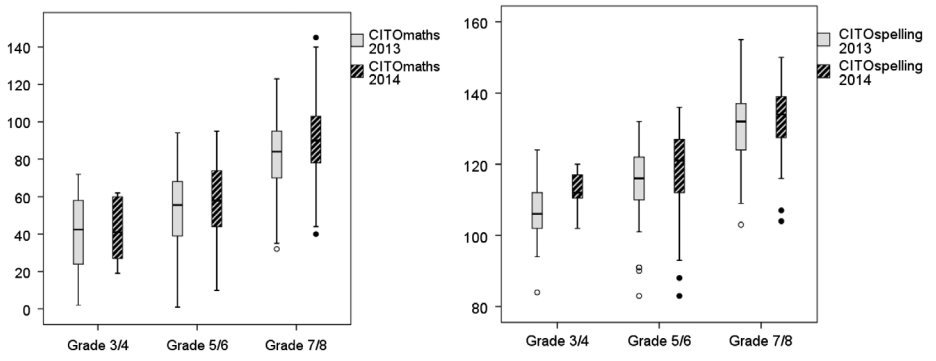

Figure 3. Students' score on $CITO_{math}$ and $CITO_{Spelling}$ in grade 3/4, 5/6 and 7/8.

Table 2. Model estimates for the variable effect on the SDQ (N = 234).

Model	Empty model	Final model
Fixed part	*Coefficient (SE)*	*Coefficient (SE)*
Intercept	15.47 (0.52)	18.35 (5.28)
PDCA		−0.99 (2.00)
AI		−0.013 (1.94)
Random part	*Variance (SE)*	*Variance (SE)*
Class level	6.080 (2.88)	5.94 (2.86)
Student level	35.524 (3.73)	35.55 (3.73)
Deviance	1529.09	1528.74

Note: No other variables contribute significantly to the model.

max = 41, SD = 12.90) and for $CITO_{Spelling}$ 3.34 (min = −19, max = 32, SD = 6.23). About 25% of the students scored negatively, i.e. their scores deteriorated instead of improving. The box plots (Figure 3) show the variation in score within the different grades (3/4, 5/6 and 7/8).

Multilevel analyses

Behaviour

In the multilevel analyses, the empty model for SDQ revealed that about 15% of the total variance in SDQ may be attributed to differences between teachers. This indicates that most differences in behaviour are due to individual student differences rather than differences between teachers. As shown in Table 2, the overall mean (intercept) of the SDQ is estimated at 15.47 (SE = 0.52). The between-teacher variance (class level) is estimated at 6.08 (2.88) and individual/residual variance as 35.52 (3.73). We then added PDCAQ and AIQ. The outcomes of the resultant models reveal no statistically significant outcome; in practice the coefficients of the added predictors are never more than twice the size of their standard error (Table 2). Comparing the variances of the added model with the variances of the empty model shows no significant differences, suggesting that the addition of the independent variables do not reveal an improvement on the empty model. Therefore, the degree of systematic (PDCA) academic (AI) instruction does not affect the outcome on the SDQ. None of the other variables (Gender, Age, IQ) contributed significantly to the model.

Table 3. Model estimates for the variable effect on the $CITO_{Maths2014}$.

Model	Empty model ($n = 200$)	Model 1 ($n = 200$)	Final model ($n = 149$)
Fixed part	*Coefficient (SE)*	*Coefficient (SE)*	*Coefficient (SE)*
Intercept	68.50 (3.41)	41.27 (19.37)	19.40 (12.81)
PDCA		16.27 (6.75)*	−4.18 (4.31)
AI		−17.71 (6.81)*	0.77 (4.33)
Grade 5/6		15.40 (5.47)*	5.17 (3.84)
Grade 7/8		45.78 (5.36)*	13.90 (4.23)*
$CITO_{Maths2013}$			0.81 (0.06)*
Random part	*Variance (SE)*	*Variance (SE)*	*Variance (SE)*
Class level	502.99 (115.82)	86.72 (32.78)	7.53 (11.01)
Student level	273.46 (31.85)	279.17 (32.03)	136.86 (18.47)
Deviance	1793.52	1733.51	1163.20

*Significant at $p < 0.005$.

Academic performance

Concerning the academic performance in maths ($CITO_{Maths2014}$), the model without predictors revealed that about 65% of the variance in score may be attributed to differences between teachers. The mean score was estimated at 68.50. The between-teacher variance is estimated at 502.99 and the residual variance as 273.46 (Table 3). After adding the explanatory variables, model 1 shows a significant effect on PDCA, AI and grades. This indicates that teachers with a higher score on PDCA, teachers with a lower score for AI and teachers in higher grades generate higher scores for $CITO_{Maths2014}$. However, the effect of PDCA and AI disappears when $CITO_{Maths2013}$ is added. The final model shows a comprehensible significant effect for grade 7/8, indicating that students in this grade score higher than students in grade 3/4. Moreover, $CITO_{Maths2013}$ is significant, indicating that students with a higher score for $CITO_{Maths2013,}$ score higher for $CITO_{Maths2014}$. Finally, after correction for grades and baseline, only 5% of the variation in score may be attributed to differences between teachers.

The model without predictors reveals that about 61% of the variance in score for the academic performance in spelling ($CITO_{Spelling}$) may be attributed to differences between teachers. The mean scores were estimated at 123.95. The between-teacher variance is estimated at 105.78 and the residual variance at 66.69. Adding the explanatory variables (model 1) shows a significant effect on PDCA and grades and no effect on AI. However, again these effects disappear when $CITO_{Spelling2013}$ is added (Table 4).Thus after correction for grades and $CITO_{Spelling2013}$, 28% of the variance in score can be attributed to the differences between teachers.

Table 4. Model estimates for the variable effect on the $CITO_{Spelling'14}$ (N = 197).

Model	Empty model ($n = 197$)	Model 1 ($n = 197$)	Final model ($n = 159$)
Fixed part	*Coefficient (SE)*	*Coefficient (SE)*	*Coefficient (SE)*
Intercept	123.95 (1.61)	98.36 (10.11)	32.48 (8.12)
PDCA		10.59 (3.45)*	−0.04 (2.22)
AI		−6.65(3.50)	0.68 (2.31)
Grade 5/6		8.16 (2.96)*	−1.19 (2.07)
Grade 7/8		19.58 (2.96)*	−0.29 (2.27)
$CITO_{Spelling2013}$			0.75 (0.05)*
Random part	*Variance (SE)*	*Variance (SE)*	*Variance (SE)*
Class level	105.78 (25.20)	21.93 (8.56)	8.84 (3.39)
Student level	66.96 (7.76)	75.35 (8.70)	22.79 (3.00)
Deviance	1482.25	1447.70	986.16

*Significant at $p < 0.005$.

No other variables (Gender, Age, IQ) contributed significantly to the model on academic performance.

Conclusion

The purpose of this study was to investigate the effect of systematic (PDCA) academic (AI) instruction on (1) behaviour and (2) academic performance of students with EBD. With respect to the first aim, it can be concluded that neither PDCA nor AI relate to students' behaviour. A higher amount of systematic academic instruction does not diminish problem behaviour. On the second aim, it can be concluded that the degree of systematic teaching (PDCA) does relate to students' academic outcomes. The higher the amount of PDCA, the higher the $CITO_{2014}$ scores. An effect was also found on academic instruction (AI), although reversed and only for maths. Thus, the higher the amount of AI, the lower the $CITO_{Maths\ 2014}$ scores. However, these effects disappears if we correct for the students' baseline scores ($CITO_{2013}$). At that point, the $CITO_{2014}$ score is mainly determined by the $CITO_{2013}$ score. Hence the outcomes of this study do not confirm the expectations based on the results of several case studies. These results are probably affected by the design of the study, namely observational rather than experimental. The following exploration of the dependent and independent variables might also enhance our understanding of the results found.

Regarding the independent variables, the relation found between the amount of systematic teaching (PDCA) and the students' score in $CITO_{2014}$ suggest that teachers preparing students for the higher echelons of CITO provide more systematic instruction. Since this effect is unchanged when we correct for grades, it is valid for teachers in all grades. It seems that higher scoring students are taught more systematically. Although it is hard to distinguish cause from effect in this case, students with a higher CITO score might necessitate a more systematic approach. However, the fact that this effect disappeared as soon as the outcomes were corrected for the baseline scores ($CITO_{2013}$) shows that a systematic approach does not expedite students' academic progress within six months. Ultimately, the $CITO_{2013}$ scores prove to be decisive for the $CITO_{2014}$ scores. This confirms Moelands' statement that the best predictor of academic outcomes of students is the preceding outcomes (2007). Regarding AI, it is difficult to explain the negative relationship with academic outcomes – the lower the AI, the higher the academic outcomes. A possible explanation could be that the better students perform, the less instruction they are given (scarcity dilemma, fide infra). After all, research shows discrepancies between academic instruction given by teachers and academic instruction received by students. Quite a high score for academic instruction by teachers does not automatically mean an equally high score for instruction received by students. Research shows that most academic instruction in classrooms with students with EBD is given individually (Vaughn et al. 2002; Van der Worp-van der Kamp et al. 2015), which possibly results in only a few students in a classroom actually being academically engaged at one time. Time spent on one student simply cannot be spent on other students. Therefore, academic instruction may be interpreted as resources (teacher efforts) that are limited in relation to demand (individual students needing instruction in order to achieve), (Gerber 2005). Gerber and Semmel (1985) refer to this as a "scarcity dilemma".

Another important issue that emerges from these findings is the reliability of the tests used for students in special education. With respect to the dependent student variables, the scores on the SDQ are noteworthy. Compared to Dutch norm scores, less than 50% of the

students fall into the clinical range. This seems to contradict the admission requirements of special education (vide supra). A possible explanation could be that teachers in special education, because they are used to problem behaviour, report more non-severe problems than teachers in general education (Van Huizen and Veerman 1999). Consequently, some teachers may provide biased reports of children's behaviour (Taylor, Gunter, and Slate 2001), possibly as a result of Posthums' Law (Van der Wolf and van Beukering 2009). The latter refers to teachers' tendency to judge a quarter of the students as fine, half as mediocre and a quarter as problematic. Therefore, the measures for behaviour may be confounded by teacher bias. Furthermore, the CITO test scores are also noteworthy. The progression in spelling and maths scores that was calculated per student showed a remarkable variety in academic outcomes. Ignoring the outliers, about 50% of the students show a considerable progression or decline in score (comparable to a progress or decline of one to several years). These wide-ranging CITO results give rise to questions about their reliability. As Fore, Boon, and Martin (2007) put it, unique characteristics of students with EBD may affect the technical adequacy of measurements. Moreover, a considerable number of students the academic progress in maths and spelling could not be measured due to overall low participation in the tests (64 and 68% respectively). This is not exceptional; George and Vannest (2009) even conclude that nearly half of the students with EBD do not participate in national reading assessments. Although we are unaware of the exact underlying reasons for not participating in assessments, it is reasonable to assume that this involves students with certain characteristics (George and Vannest 2009). Consequently, data are probably not missing randomly but for specific reasons. Moreover, since valid assessments are an important aspect of systematic work, their scarcity could also be a reason for this study's disappointing results.

Some limitations of this study need to be acknowledged. Although the multilevel models include multiple explanatory variables that could predict the effect of behavioural and academic outcomes, the model possibly misses important additional factors that predict students' outcomes. Therefore, the results may still be affected by hidden (i.e. an omitted variable) bias. For example, the period of time students received special education was not included. Moreover, because the design was not randomised, the possibility exists that the previous teacher may also have played a role in establishing the outcomes. In addition the study was limited to the effect within six months of teaching. Analyses over a longer period of time may have yielded stronger effects.

Although the results of this study did not confirm our expectations, they are significant in two major aspects. First, teachers see themselves as working fairly systematically and giving an adequate amount of academic instruction. These outcomes were higher than could be expected from the literature. Therefore, the outcomes indicate a shift in focus from ad hoc to systematic and from behavioural to academic instruction. The continuing claims of educational researchers that many teachers work in an ad hoc way while focusing on behaviour are not confirmed by the outcomes of this study. Recent government efforts to improve the learning outcomes of all students including those with special needs are probably starting to be effective. In particular quality indicators from the Dutch Inspection Framework, which strongly emphasise cognitive outcomes in basic subjects, are increasingly evident in special education. However, the realisation of systematic academic instruction in daily practice needs further research. The combination of findings from this study suggests that some important conditions for being successful in systematic academic instruction are not yet adequately met. For instance, systematic instruction requires reliable assessments.

Shriner et al. (2014) even refer to assessment as the keystone of special educational programming. Yet, the participation in and suitability of the national assessments for students with EBD is questionable. Since the Dutch Inspectorate relies heavily on student performance in the CITOs, it is important to investigate the reliability of national assessments for students with EBD. Another point of particular interest is the difference between given and received instruction. Besides insight into the amount of instruction given, it seems equally important to focus on the amount of instruction each student receives during lessons. Future studies on these two topics are therefore recommended.

Returning to the question posed at the beginning of this study, the findings underline the risk of translating effects shown in controlled small-scale studies to the daily practice of special education (Morgan et al. 2010). It is possible that the necessary amount of systematic academic instruction can only be achieved on a desirable scale in small case studies. Several reviews about academic interventions regarding students with EBD show that numerous studies examine the effects of interventions on a small number of participants, with researchers or their assistants as the primary interventionists (Therrien et al. 2014; van der Worp-van der Kamp et al. 2014). This is not comparable to this study's participating schools with classes of two or three times as many students and with only one teacher. Gerber speaks in this context of "well-validated research findings, difficult to implement at scale" (2005, 522). Therefore, notwithstanding the increasing interest in educating students with EBD, a lot of research is still needed on optimising it. Gaining insight into the unruly daily practice of special education seems to be an important aspect of this.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Lidy van der Worp-van der kamp is a Phd in Special Educational Needs. Presently, she is working as a lecturer and researcher at the Department of Special Needs Education, University of Groningen, The Netherlands. Her research focuses on the effect of academic instruction on academic and behavioural outcomes of students in special education.

Sip Jan Pijl is a professor of Special Needs Education and education director at the University of Groningen, The Netherlands. He is also attached part-time to the Norwegian University of Science and Technology in Trondheim, Norway. His research focuses on both conditions and effects of inclusive education.

Jan Bijstra is a doctor in Developmental Psychology. Presently, he is working as a researcher at the Regional Expertise Centre for Northern Netherlands (RENN4) in the Netherlands. His research interest includes the effects of intervention programmes on students with special educational needs.

Wendy Post is an associate professor in statistics and methodology at the Department of Special Needs Education, University of Groningen, The Netherlands. She is specialized in random effects models, missing data and measurement theory.

Els van den Bosch has worked for nearly 25 years as lecturer at the Department of Special Needs Education, University of Groningen, The Netherlands . She had a broad interest in students with special educational needs, including gifted students. Sadly, she died just before the completion of this article.

References

Becker, A., W. Woerner, M. Hasselhorn, T. Banaschewski, and A. Rothenberger. 2004. "Validation of the Parent and Teacher SDQ in a Clinical Sample." *European Child and Adolescent Psychiatry* 13 (2): 11–16.

Bland, J. M., and D. G. Altman. 1986. "Statistical Methods for Assessing Agreement between Two Methods of Clinical Measurement." *Lancet* 327 (8476): 307–310.

Bos, C. S., and S. Vaughn. 2006. *Strategies for Teaching Students with Learning and Behavior Problems*. Boston, MA: Pearson, Allyn, and Bacon.

Bradley, R., J. Doolittle, and R. Bartolotta. 2008. "Building on the Data and Adding to the Discussion: The Experiences and Outcomes of Students with Emotional Disturbance." *Journal of Behavioral Education.* 17 (1): 4–23.

Coleman, M., and S. Vaughn. 2000. "Reading Interventions for Students with Emotional/Behavioral Disorders." *Behavioral Disorders* 25 (2): 93–104.

Fore, C., R. Boon, and C. Martin. 2007. "Concurrent and Predictive Criterion-related Validity of Curriculum-based Measurement for Students with Emotional and Behavioral Disorders." *International Journal of Special Education* 22 (2): 24–31.

George, C. C., and K. J. Vannest. 2009. "To What Extent Do Students with EBD Participate in State Accountability Assessments? Implications for Teachers." *Beyond Behavior* 18 (3): 33–39.

Gerber, M. M. 2005. "Teachers are Still the Test: Limitations of Response to Instruction Strategies for Identifying Children with Learning Disabilities." *Journal of Learning Disabilities* 38 (6): 516–524.

Gerber, M. M., and M. I. Semmel. 1985. "The Microeconomics of Referral and Reintegration: A Paradigm for Evaluation of Special Education." *Studies in Educational Evaluation* 11 (1): 13–29.

Hagaman, J. L. 2012. "Academic Instruction and Students with Emotional and Behavioral Disorders." In *Behavioral Disorders: Practice Concerns and Students with EBD (Advances in Special Education)*, edited by J. P. Bakken, F. E. Obiakor, and A. R. Rotatori, 23–41. Bingley: Emerald Group Publishing.

Kauffman, J. M., and T. J. Landrum. 2013. *Characteristics of Emotional and Behavioral Disorders of Children and Youth*. 10th ed. Upper Saddle River, NJ: Merrill Prentice-Hall.

Koopman, P. N. J., and G. Ledoux. 2013. *Kengetallen Passend Onderwijs*. Amsterdam: Kohnstamm Instituut.

Lane, K. L., J. H. Wehby, M. A. Little, and C. Cooley. 2005. "Academic, Social and Behavioral Profiles of Students with Emotional and Behavioral Disorders Educated in Self-contained Classrooms and Self-contained Schools: Part I – Are they more Alike than Different?" *Behavior Disorders* 30: 349–361.

Ledoux, G., J. Roeleveld, A. Van Langen, and E. Smeets. 2012. *COOL Speciaal. Inhoudelijk rapport* [*COOL Special. Explanatory Report*]. Amsterdam: Kohnstamm Instituut, Nijmegen: ITS.

Lee, Y., G. Sugai, and R. Horner. 1999. "Effect of Component Skill Instruction on Math Performance and On-task, Problem, and Off-task Behavior of Students with Emotional and Behavioral Disorders." *Journal of Positive Behavior Interventions* 1: 195–204.

Levy, S., and D. J. Chard. 2001. "Research on Reading Instruction for Students with Emotional and Behavioural Disorders." *International Journal of Disability, Development, and Education* 48: 429–444.

Levy, S., and S. Vaughn. 2002. "An Observational Study of Teachers' Reading Instruction of Students with Emotional or Behavioral Disorders." *Behavioral Disorders* 27 (3): 215–235.

Moelands, A. 2007. *Stand van Zaken Leerrendementsverwachting* [*State of Affairs Expected Learning Output*]. Arnhem: CITO.

Mooney, P., M. H. Epstein, R. Reid, and J. R. Nelson. 2003. "Status of and Trends in Academic Intervention in Research for Students with Emotional Disturbance." *Remedial and Special Education* 24 (5): 273–287.

Mooney, P., J. B. Ryan, B. M. Uhing, R. Reid, and M. H. Epstein. 2005. "A Review of Self-Management Interventions Targeting Academic Outcomes for Students with Emotional and Behavioral Disorders." *Journal of Behavioral Education* 14 (3): 203–221.

Morgan, P. L., M. Frisco, G. Farkas, and J. Hibel. 2010. "A Propensity Score Matching Analysis of the Effects of Special Education Services." *The Journal of Special Education* 43 (4): 236–254.

Nelson, J. R., G. J. Benner, K. L. Lane, and B. W. Smith. 2004. "Academic Achievement of K-12 Students with Emotional and Behavioral Disorders." *Exceptional Children* 71: 59–73.

Nelson, J. R., G. J. Benner, and P. Mooney. 2008. *Instructional Practices for Students with Behavioral Disorders: Strategies for Reading, Writing, and Math*. New York, NY: Gilford Press.

Pianta, R. C., and B. K. Hamre. 2009. "Conceptualization, Measurement, and Improvement of Classroom Processes: Standardized Observation can Leverage Capacity." *Educational Researcher* 38 (2): 109–119.

Rasbash, J., F. Steele, W. Browne, and B. Prosser. 2005. *A User's Guide to MLwiN*. Bristol: Centre for Multilevel Modeling.

Reid, R., J. Gonzalez, P. D. Nordness, A. Trout, and M. H. Epstein. 2004. "A Meta-analysis of the Academic Status of Students with Emotional/Behavioral Disturbance." *The Journal of Special Education* 38 (3): 130–143.

Sacks, G., and L. Kern. 2008. "A Comparison of Quality of Life Variables for Students with Emotional and Behavioral Disorders and Students without Disabilities." *Journal of Behavioral Education* 17 (1): 111–127.

Shriner, J. G., S. P. Ardoin, M. L. Yell, and S. J. Carty. 2014. Assessment of Students with Emotional and Behavioral Disorders. In *Evidence Based Practices for Educating Students with Emotional and Behavioral Disorders* (2nd ed.), edited by M. L. Yell, N. Meadows, E. Drasgow, and J. Shriner, 45–60. Upper Saddle River, NJ: Pearson, Merrill Education.

Taylor, P. B., P. L. Gunter, and J. R. Slate. 2001. "Teachers' Perceptions of Inappropriate Student Behavior as a Function of Teachers' and Students' Gender and Ethnic Background." *Behavioral Disorders* 26: 146–151.

Therrien, W. J., J. C. Taylor, S. Watt, and E. R. Kaldenberg. 2014. "Science Instruction for Students with Emotional and Behavioral Disorders." *Remedial and Special Education* 35 (1): 15–27.

Umbreit, J., J. Ferro, C. Liaupsin, and K. Lane. 2007. *Functional Behavioral Assessment and Function-based Interventions: An Effective, Practical Approach*. Upper Saddle River, NJ: Prentice-Hall.

Van der Wolf, K., and T. van Beukering. 2009. *Gedragsproblemen in Scholen. Het Denken en Handelen van Leraren*. Leuven: Acco.

Van der Worp-van der Kamp, L., S. J. Pijl, J. O. Bijstra, W. J. Post, and E. J. van den Bosch. 2015. "Systematic Academic Instruction for Students with EBD: The Construction and Use of a Tool for Teachers." *Journal of Research in Special Educational Needs*. Advance online publication. doi: http://dx.doi.org/10.1111/1471-3802.12096.

Van Huizen, J. H., and J. W. Veerman. 1999. "Teacher Ratings of Children's Emotional and Behavioral Problems in a Dutch Special Education Sample." *Pedagogische Studiën* 76 (5): 320–331.

Vannest, K., K. Temple-Harvey, and B. Mason. 2009. "Adequate Yearly Progress for Students with Emotional and Behavioral Disorders through Research-based Practices." *Preventing School Failure: Alternative Education for Children and Youth* 53 (2): 73–84.

Vaughn, S., S. Levy, M. Coleman, and C. S. Bos. 2002. "Reading Instruction for Students with LD and EBD: A Synthesis of Observation Studies." *The Journal of Special Education* 36 (1): 2–13.

WEC-raad. 2008. *Wijzigingen Indicatiecriteria Voor Leerlingebonden Financiering* [*Changing Accession Criteria Student-bound Budget*]. http://www.projectenso.nl/docs/10b31bf0-d4c4-8c8b-e8d8-624a2bf22c44.pdf.

van Widenfelt, B. M., A. W. Goedhart, P. D. A. Treffers, and R. Goodman. 2003. "Dutch Version of the Strengths and Difficulties Questionnaire (SDQ)." *European Child and Adolescent Psychiatry* 12 (6): 281–289.

Wilen, W., J. Hutchinson, and M. Ishler. 2008. *Dynamics of Effective Secondary Teaching*. 6th ed. Boston, MA: Pearson.

van der Worp-van der Kamp, L., S. J. Pijl, J. O. Bijstra, E. J. van den Bosch, and E. J. van den Bosch. 2014. "Teaching Academic Skills as an Answer to Behavioural Problems of Students with Emotional or Behavioural Disorders: A Review." *European Journal of Special Needs Education* 29 (1): 29–46.

Yell M. L., T. W. Busch, and D. C. Rogers. 2014. "Teaching Students with EBD III: Planning Instruction and Collecting Data to Monitor Student Progress." In *Evidence Based Practices for Educating Students with Emotional and Behavioral Disorders* (2nd ed.), edited by M. L. Yell, N. Meadows, E. Drasgow, and J. Shriner. Upper Saddle River, NJ: Pearson, Merrill Education.