

# The psychometric evaluation of a questionnaire to measure attitudes towards inclusive education

Anke de Boer · Marieke Timmerman · Sip Jan Pijl · Alexander Minnaert

Received: 10 May 2011 / Revised: 8 September 2011 / Accepted: 15 December 2011 /  
Published online: 6 January 2012  
© The Author(s) 2012. This article is published with open access at Springerlink.com

**Abstract** In many countries, there has been a development towards the inclusion of students with special educational needs in regular education. Over the past decade, this has resulted in an increased interest in attitudes towards this educational change of those directly involved. This current study aims at the development, psychometric evaluation, and validation of a questionnaire to measure attitudes of teachers, parents, and students. The questionnaire is based on the three-component theory, reflecting a cognitive, an affective, and a behavioral component. Based on data of a pilot study and a main study, we evaluated the psychometric properties and separability of the three components per questionnaire using a Mokken scale analysis. Removing various dysfunctional items resulted in questionnaires with appropriate psychometric properties and high reliability. Results of the analysis for the separability of the attitude components revealed no distinction could be made between the components in the teachers' and parents' questionnaire. In the students' questionnaire, the items belonging to the cognitive and affective/behavioral components were distinguished as two scales.

**Keywords** Teachers · Parents · Students · Attitude questionnaire · Mokken scale analysis · Special educational needs · Inclusion

## Introduction

The inclusion of children with disabilities in regular education became an important matter of debate in the last decades. In many countries, this resulted in changing education policies and in making regular education the standard provision for all students, including those with special educational needs. One of the core ideas behind these changes is that both children with and without disabilities experience social benefits in regular schools (Flem and Keller 2000), like acceptance of children with disabilities in society (Tafa and Manolitsis 2003) and friendships (Scheepstra et al. 1999). Although more acceptance and friendships for students

---

A. de Boer (✉) · M. Timmerman · S. J. Pijl · A. Minnaert  
University of Groningen, Groningen, Netherlands  
e-mail: anke.de.boer@rug.nl

with disabilities in regular schools are a widely supported aim, implementing inclusion in daily practice is challenging (Ferguson 2008; Vislie 2003).

The difficulties in making inclusive education happen triggered researchers to think about the reasons for this. Several authors point at the key role that teachers play in realising inclusive education (Meijer 2003; Norwich 1994). Their attitudes are regarded as highly important (Chow and Winzer 1992), but those of others directly involved, like parents and students, are of relevance as well. According to Stoiber et al. (1998), the voices of those directly involved in change should be heard because they provide valuable “inside” perceptions and information.

To obtain more information about these voices, educational research has increasingly focused on measuring attitudes towards inclusive education. This resulted in the construction and evaluation of various scales to measure attitudes of teachers, parents, and students towards various aspects of inclusion (Vignes et al. 2008). Some of those scales use the widely known three-component theory (Eagly and Chaiken 1993; Triandis 1971) as a theoretical framework. According to this, attitudes are considered to comprise three components: (1) cognitive, (2) affective, and (3) behavioral. The first reflects ones beliefs and knowledge about the subject (i.e., children with special needs), the second reflects a person’s feelings, and the third component reflects ones behavioral intentions.

In the field of attitude research, some argue in favour of this three-component model (e.g., Triandis 1971), while others prefer to work with a two (e.g., Ajzen 2005) or a single-component model (e.g., Dillon and Kumar 1985). The three-component model assumes that the three attitude components are separate constructs (Eagly and Chaiken 1993; Ostrom 1969). The two-component one usually distinguishes the cognitive and affective components, while the behavioral intentions are excluded (Bagozzi and Burnkrant 1979, 1985; Fishbein and Ajzen 1974). The single-component model, however, assumes that a distinction between the three components cannot be sensibly made (Dillon and Kumar 1985). The strong relationship between the three components is also underlined by Albarracín et al. (2005). They state that attitudes are evaluative tendencies, which can both be inferred from, and have an influence on, beliefs, affect, and behavior: “beliefs, affect and behaviour are seen as interacting with attitudes rather than as being their parts” (p. 5).

A number of researchers have tried to confirm the discriminate validity of the three attitude components. Some of the results supported the three-component model (Breckler 1984; Mahat 2008; Rosenbaum et al. 1986), while others established a two- or a single-component model (Bagozzi and Burnkrant 1985; Fishbein and Ajzen 1974). For example, Ajzen (2005) stated that most data reported in literature are quite consistent with a single-component model because factor analyses revealed a single factor explaining most of the variance present in the data. Eagly and Chaiken (1993), however, underline that evidence supports the separability of the three components under certain circumstances. Thus, the number of attitude components is a still a matter of debate.

Although views may differ in this discussion, it is evident that attitude measurements should be based on a well-considered, conceptual framework. A closer look at the conceptual framework behind a number of attitude scales revealed that many of these lack any theoretical basis (De Boer et al. 2010; De Boer, Pijl & Minnaert 2011), which hampers the interpretation of the scales.

The purpose of the current study is to develop an instrument to measure attitudes of Dutch teachers, parents, and students towards inclusive education. This self-completion questionnaire for each of the three target groups includes all three attitude components. The current study examines the psychometric properties of the teacher, parent, and student questionnaires and addresses the empirical support for the three attitude components.

## Method

### Procedure for questionnaire development

The procedure used to develop the questionnaire consisted of four steps. First, based on content analysis, a set of items was proposed for each subscale for each target group. Second, the quality of the scales was assessed in a pilot study. Then, the questionnaires were adapted based on the results of the study. Finally, the revised version was examined in the main study. The four steps will now be described in greater detail.

### Conceptual framework of the questionnaire

Three review studies were performed to examine what attitudes teachers, parents, and students hold towards the inclusion of children with special needs. The studies selected for the reviews were derived from an electronic search in EBSCOhost Complete. The selection was based on the following criteria: The study was published between 1998 and 2008 and focused on attitudes of regular primary school teachers, parents, or students towards the inclusion of children with special needs in regular primary schools.

The total number of studies included in the three review studies was 58. In order to describe the outcomes of the studies, these were grouped according to the three target populations (i.e., teachers, parents, and students) and according to the three attitude components. This grouping was based on analyzing the components measured by the (sub)scales. In some cases, it was possible to categorize (sub)scales as belonging to one of the three components. In others, only careful analysis of the type of questions or statements used in the questionnaires revealed on which component(s) of attitude the questionnaire was focused. These analyses showed that only a few studies used the three-component theory as a framework and defined their questionnaires in terms of cognitive, affective, or behavioral aspects (Alderfer et al. 2001; Beck et al. 2006; Nowicki 2006).

The few studies in which (sub)scales were used for a particular component gave an insight into the scale developers' views on the content of a particular attitude component. For example, the teacher scales designed to measure the cognitive component largely included items reflecting general beliefs about the philosophy behind inclusion, beliefs on children's educational rights, and knowledge about disabilities. For this group, the affective component comprised items reflecting feelings like competence, confidence, frustration, and irritation, while items belonging to the behavioral component reflected teachers' willingness to implement educational change (e.g., change the curriculum).

The scales to measure the cognitive component of parents included mainly items about general beliefs on the philosophy of inclusion, special needs children's educational rights, and knowledge about disabilities. The affective component was frequently measured by items reflecting concerns, fears, and approval (e.g., allowing their child to play with a disabled child), while the behavioral component was measured by items on their willingness to interact and support (e.g., personally standing up for a child with a disability who's bullied).

The cognitive component in the student scales often focused on items reflecting their knowledge about the behavior of children with disabilities. Feelings of fear, shame, and joy were measured by items belonging to the affective component. The behavioral component was often measured by items reflecting students' school and free time related willingness to interact and support.

The content analysis of the (sub)scales resulted in a list of concepts intended for use in the teacher, parent, and student questionnaires. The goal was to find existing teacher, parent, and student scales preferably comprising all three attitude components and each with well-documented psychometric quality. In case this was not possible, existing subscales, or parts of subscales, were used to build new scales measuring all components. The new scales were then translated into Dutch and adjusted to the Dutch situation.

### Selection of (sub)scales

The selection of parts of the (sub)scales was based on the following criteria:

1. Reflecting relevant constructs of the cognitive, affective and behavioral component;
2. The reliability of the original scales in the relevant population was reported to be  $>0.80$ , which is considered to be appropriate for research purposes (Drenth and Sijtsma 2006).

*Teacher questionnaire* The cognitive component was based on 12 items of the subscale “core perspectives” from the “My Thinking About Inclusion” questionnaire (MTAI; developed by Stoiber et al. 1998). For the affective component, four items of the “Multidimensional Attitudes Towards Inclusive Education Scale” (MATIES; developed by Mahat 2008), were selected to measure feelings of frustration and irritation. Because those items only reflected negative feelings, eight items from the “Skills” questionnaire (developed by Avramidis et al. 2000) were added. These reflected positive feelings, like competence and confidence. The subscale measuring the behavioral component was based on six items from the MATIES (Mahat 2008), reflecting teachers’ willingness to interact and support.

A total number of 30 items for the teachers’ questionnaire were selected (see Table 1). Those items pertained to the three components with respectively 12, 12, and six items measuring the cognitive, affective, and behavioral domains. Respondents could rate the items on a 5-point Likert scale (1=completely disagree to 5=completely agree).

*Parent questionnaire* The “Parental Attitudes Toward Handicapped Children” questionnaire (PATHC; developed by Rosenbaum et al. 1987) was used as ‘donor’ scale. The questionnaire consists of items reflecting all three attitude components, but the concept “beliefs of parents towards inclusion” was insufficiently covered. Thus, eight items of the subscale “core perspectives” from the “My Thinking About Inclusion” questionnaire (Stoiber et al. 1998) were added.

**Table 1** Overview of selected questionnaires

Target group	Instrument	Items per component <sup>b</sup>			References	Reliability
		C	A	B		
Teachers	MTAI—core perspectives	12			Stoiber et al. (1998)	Alpha=0.80
	Skills		8		Avramidis et al. (2000)	Alpha=0.88
	MATIES		4	6	Mahat (2008)	Alpha=0.91
Parents	PATHC	11	11	8	Rosenbaum et al. (1987)	Alpha=0.89
	MTAI—core perspectives	8			Stoiber et al. (1998)	Alpha=0.80
Students	CATCH	12	12	12	Rosenbaum et al. (1986)	Alpha=0.90

<sup>a</sup> The theoretical framework of the questionnaire is based on the three-component theory

<sup>b</sup> Attitude component: C cognitive, A affective, B behavioral

A total number of 38 items for parents' questionnaire were selected (see Table 1) relating to the three components, with 19 (cognitive), 11 (affective), and eight (behavior) items, respectively. Respondents could rate the items on a 5-point Likert scale (1=completely disagree to 5=completely agree).

*Student questionnaire* The “Chedoke-McMaster Attitudes Towards Children with Handicaps” (CATCH; developed by Rosenbaum et al. 1986) included all three components and met our selection criteria. Hence, it was decided to select this questionnaire to measure students' attitudes. This questionnaire consisted of 36 items, 12 items for each component (see Table 1).

The original CATCH questionnaire used the term “handicapped child.” It is likely that students aged 8–12 years would interpret this in many different ways. To avoid this, vignettes were presented at the beginning of the questionnaire. These vignettes represented a hypothetical child showing characteristics of a specific type of disability, but without using terms as handicapped, disability, impairment, or special needs (see Appendix 1). Students were asked to read this story before responding to the questionnaire. The wording of the items in the student questionnaire was linked to the content of the vignette. Three different vignettes were compiled describing children, which could attend a regular or a special Dutch school. The stories were compiled by the first author and verified by an educational psychologist.

The students' questionnaire consisted of a vignette and 36 statements, with each component being measured by 12 items. The students' questionnaire had three versions with each having a different vignette. Respondents could rate the items on a 5-point Likert scale (1=completely disagree to 5=completely agree).

#### Translation procedure

The selected items for the three scales were translated in order to assess the attitudes of Dutch teachers, parents, and students: Two people (the first author and a graduate student) translated the items independently of each other. Both translations were then compared, and both translators discussed what the most accurate translation would be.

After the translation of the items, they were tested in a small convenience sample. This made it possible to detect difficulties in the wording of the items. From each target population, a small group (seven teachers, two parents, and four students) were asked to complete the questionnaire with the graduate student or the first author present. All remarks by the teachers, parents, and students on the formulation of the items were examined. Subsequently, the wording of difficult or vague formulated items was changed.

#### Procedure for questionnaire evaluation

The psychometric properties and the construct validity of the questionnaires was tested in a two-stage process, the first stage comprising drawing three independent samples of teachers ( $N=58$ ), parents ( $N=58$ ), and students ( $N=1,157$ ). These data were used for a first analysis of the psychometric properties of the scales and the separability of the three attitude components in each scale. Based on the outcomes of these first analyses, adaptations and improvements to the questionnaires could be made. The second stage was set up in order to examine the psychometric properties and the possible separability of the attitude components of the three final scales.

## Analysis of the psychometric properties

The analysis of the psychometric properties was based on an item response theory model. To analyze the item quality of each questionnaire a Mokken scale analysis (MSA) was used. In a MSA, the relationship between observed item responses and a latent trait are examined. The latent trait is assumed to pertain to the measured concept (e.g., attitudes). Based on MSA, the quality of individual items can be assessed. In particular, the Monotonicity Homogeneity Model was used (MHM; Mokken 1971; Sijtsma and Molenaar 2002) based on the following three assumptions:

1. Unidimensionality of items: a single latent trait underlies the items responses
2. Local independence: the scores on a previous item are not to be influenced by the scores on the other items of the scale
3. Monotonicity: all item response functions (IRFs), which express the relationship between the latent trait and the item scores, are monotone, non-decreasing.

When those three assumptions are met for a set of items, the important implication is that the individuals' sum scores of the items provide the ordering of the individuals on the latent trait.

The program "Mokken Scale Analysis for Polytomous" items (Molenaar and Sijtsma 2000) was used to analyze the fit of the MHM for each subscale (i.e., attitude component). In this way, it can be determined whether the sum scores of the items indeed indicate the ordering of persons on, e.g., the cognitive component, or whether items should be adapted or deleted from the scale before the scale can sensibly use to order the persons. In order to assess whether item  $i$  fits the MHM, the item scalability coefficient ( $H_i$ ) of each item was examined. If the MHM holds,  $H_i$  is between 0 and 1 at the population level. The  $H_i$  values furthermore indicate the extent of discrimination power of the items across individuals, with lower values indicating a lower discrimination power. Molenaar and Sijtsma (2000) suggest as rule of thumb that items associated with  $H_i$  values  $< 0.30$  are considered to have weak discrimination power. Thus, the  $H_i$  value of item  $i$  expresses the quality of the item, given the other items in the scale. The quality of a (sub)scale is indicated with the  $H$  value, indicating its strength, with  $0.30 \leq H \leq 0.40$  regarded as a weak scale,  $0.40 \leq H \leq 0.50$  a moderate scale, and an  $H > 0.50$  a strong scale.

Besides the scalability coefficients, other diagnostics were used to assess whether the monotonicity assumption holds for each item. The diagnostics are summarized into so-called criteria values, where a value  $\geq 80$  strongly suggests assumption violation, values between 40 and 80 are questionable, and values of  $\leq 40$  are satisfying (Sijtsma and Molenaar 2002). Based on the outcomes of the analysis and the content of the items, we decided whether suspicious items would be removed or maintained in the questionnaire with a reformulation.

Next to the analysis with the MHM, we examined if evidence for differential item functioning (DIF) could be found within each questionnaire per target group. DIF might occur when the IRF of a particular item is different in two relevant subgroups (i.e., men/boys and women/girls). For each of the target groups, DIF was examined among the three subgroup variables: gender, experience with inclusion/having a child with SEN, and type of vignette.

The following analysis procedure was used for each subscale:

1. Assessing the  $H_i$  coefficients and the criteria values to examine whether the assumptions of the MHM appear to hold
2. Inspection of suspicious items from (1) based on  $H_i$  and content

3. Removal of items from the subscales/adaptations of the content of the item
4. Assessing the diagnostics for differential item functioning across subgroups
5. Determination of the  $H$  value of the final scale (thus, excluding suspicious items) and the reliability coefficients.

#### Analysis of the separability of the subscales

To determine whether the attitude components can be seen as separate subscales within the attitude questionnaire, we used the automatic item step procedure (AISP) of MSA. This procedure can be used to select a cluster of items from a larger set, where each cluster consists of items that measure the same latent trait, with sufficient discrimination power. AISP can be seen as an alternative to factor analysis and applies very well to items from questionnaires that are scored polytomously. The AISP requires a sample of at least 100 respondents (Sijtsma and Molenaar 2002). The AISP aims at finding scales from a set of items such that each scale satisfies a minimal, prespecified degree of quality—as indicated by the scale  $H$  coefficient—and contains as much items as possible that are indicative of the same latent trait. The AISP was used according to the guidelines of Sijtsma and Molenaar (2002, p. 80–82) in which we used the bottom-up strategy with lower bounds ranging from  $c=0.30$  to  $c=0.60$ . The AISP was intended to be performed on all items using various boundary levels ( $c=0.30$  to  $c=0.60$ , decreasing each step with 0.10). With increasing boundary levels, a typical pattern in the emergence of indicated scales arise. A small boundary value results in most items in one scale. The noticeable difference between the single scale and multiple subscales occur with increasing boundary values: The emergence of one smaller scale indicates a single scale suffices, where two or more scales indicate the necessity of subscales.

## Results

### First stage analyses: procedure and participants

To evaluate the psychometric properties of the items and the separability of the subscales, an independent study per target group was performed. This means that the schools where the teacher and student studies were performed were independent of each other, while parents and students were not related as well. All three studies were conducted by undergraduates and supervised by the first author. The teacher and student studies took place in April 2009 and parents' study took place in June 2009. A description of the data collection is given below. Detailed information about the participants is presented in Table 2.

*Teachers* A total number of 60 regular primary schools were randomly selected from a list of addresses of schools situated in an urban area in the north of The Netherlands. Teachers from grades 5 to 8 were invited to participate in the study. A survey package was sent to each school containing an invitation to participate in the study, four questionnaires and return envelopes. In total, 58 questionnaires were returned, reflecting a response rate of 24%.

*Parents* The sample of parents testing the parent questionnaire was gleaned by the first author of this article inviting acquaintances with primary school age children to participate in the study. Since there was not any contact with schools at the time of the data collection, it was impossible to contact parents via teachers or students. Therefore, we decided to use a

**Table 2** Teacher and parent demographics of the pilot study and the main study

Demographics <sup>a</sup>		Pilot study		Main study	
		Teachers (N=58) N (%)	Parents (N=58) N (%)	Teachers (N=45) N (%)	Parents (N=420) N (%)
Gender	Male	13 (23)	17 (29)	11 (24)	81 (20)
	Female	45 (77)	41 (71)	34 (76)	339 (80)
Age <sup>a</sup>	21–30 years	12 (22)	5 (8)	26 (59)	27 (6)
	31–40 years	9 (15)	23 (39)	4 (8)	301 (72)
	>40 years	37 (63)	30 (53)	15 (33)	90 (21)
Teaching experience	1–4 years	8 (13)	–	12 (27)	–
	5–9 years	13 (23)	–	15 (33)	–
	10–14 years	8 (15)	–	3 (7)	–
	> 14 years	29 (49)	–	15 (33)	–
Education degree <sup>a</sup>	Primary	–	1 (2)	–	2 (1)
	Secondary	–	4 (7)	–	37 (9)
	Intermediate vocational	–	22 (38)	–	188 (45)
	Bachelor degree	–	20 (34)	–	158 (38)
	Master's degree	–	11 (19)	–	31 (7)
Having a child with SEN (in class) <sup>b</sup>	Yes	30 (52)	12 (21)	45 (100) <sup>c</sup>	132 (32)
	No	28 (48)	46 (79)	–	288 (68)

<sup>a</sup> Due to missing values of demographics the number sometimes does not correspond with the sample size

<sup>b</sup> SEN Special educational needs

<sup>c</sup> Due to other purposes of the data-collection the inclusion criterion of the second stage sampling was the inclusion of at least one child with SEN in teachers' class

snowball effect in sampling parents, i.e., invited parents who indicated they wanted to participate in the study then approached other parents and asked if they wanted to participate. Parents who wanted to participate were informed about the research and received a survey package, including an information letter, two questionnaires (one for each parent) and a return envelope. This procedure resulted in a total of 120 parents being invited to participate in the study. A total of 58 parents returned the questionnaire, reflecting a response rate of 48%.

**Students** A total of 26 schools were randomly selected from a local address list of schools situated in the north of The Netherlands. Grades 5–8 of each school (age range, 8–12) were invited by letter to participate in the study. Of these, 14 schools wished to participate (54%). In most cases, the questionnaires were personally administered by the graduate students, although in some cases, the schools only wanted to participate if teachers themselves could administer them. In such cases, the schools received written standardized instruction on how the questionnaire was to be administered.

The data of a sample of 1,157 regular primary school pupils was analyzed, consisting of 620 girls and 537 boys. Each class was randomly divided into three groups, each with a different vignette and each filling in the appropriate questionnaire. The age of the pupils ranged between 8 and 12 years old, with a mean of 10.1 years (SD=1.3). A total number of 15 children had been formally assessed as having a disability (1.2%).



## First stage analyses: evaluation of the psychometric properties

*Teacher questionnaire* The assumptions of the MHM were checked using the scalability coefficients  $H_i$  and the criteria values. Four out of 30 items had insufficient scalability coefficients (i.e.,  $H_i < 0.30$ ). Based on the outcomes and content analysis, it was decided to remove three items (e.g., “Parents of students with special needs benefit from inclusive education”). We decided to maintain the other item and adapt the wording since this was regarded as being highly relevant to the scale. Of the remaining 26 items, no violation of the monotonicity assumption was indicated. We found that the items of each component had weak to moderate scalability coefficients (see Table 3 for a summary).

Comparing the ordering of the response categories for all the statements across “gender of the teacher” showed a slight indication of DIF within the cognitive component. For two statements (e.g., “Rights of regular education for children with special needs”), it is more likely that female teachers show positive scores. Because this difference appeared to be rather small, we decided to maintain the statements in the subscale. No indication of DIF was found for the items belonging to the affective and behavioral components for “gender” and “with/without experience.”

*Parent questionnaire* The outcomes of the analysis revealed that several items violated the assumptions of the MHM (see Table 3). Based on these outcomes, it was decided to delete six of the 38 items. For five items with low scalability coefficients ( $H_i < 0.30$ ), it was decided to maintain them in the scale as their content was considered to be important to address in the scale (e.g., “A regular education teacher cannot address adequately the individual needs of children with special needs”). To obtain stronger discrimination power we changed the formulation of the items. For the other 27 items, no indication of a violation of the monotonicity assumption was found. The items of each component had weak to moderate scalability coefficients and a scale  $H$  of moderate strength.

Differential item functioning appeared to be absent for gender, while it was found for the variable “having/not having a child with a disability.” The results showed that parents having a child with a disability are more likely to show positive answers on several items of the cognitive and behavioral subscale. The differences appeared to be small. Therefore, we decided to keep the items in the subscale.

**Table 3** Summary of number of deleted items, adapted items, final scale, the scalability  $H_i$ , and scale  $H$

Group	Scale <sup>a</sup>	Deleted $H_i < 0.30$	Adapted $H_i < 0.30$	Final scale	Scalability $H_i$	Scale $H^b$
Teachers	C	2	1	10	$0.30 \leq H_i \leq 0.43$	0.35
	A	1	0	11	$0.31 \leq H_i \leq 0.50$	0.38
	B	0	0	6	$0.41 \leq H_i \leq 0.66$	0.57
Parents	C	4	2	15	$0.26 \leq H_i \leq 0.51$	0.42
	A	0	2	11	$0.34 \leq H_i \leq 0.54$	0.42
	B	2	1	6	$0.26 \leq H_i \leq 0.56$	0.42
Students	C	5	3	7	$0.26 \leq H_i \leq 0.39$	0.34
	A	4	0	8	$0.42 \leq H_i \leq 0.56$	0.50
	B	3	0	9	$0.34 \leq H_i \leq 0.55$	0.48

<sup>a</sup> C cognitive, A affective, B behavioral

<sup>b</sup> The scale  $H$  is based on the items with  $H_i > 0.30$

*Student questionnaire* The results of the analysis revealed that 12 out of 36 items had insufficient scalability coefficients, and these were removed from the scale. For three other items with low scalability coefficients, it was decided to keep them in the scale, but adapt the formulation of the items (e.g., “I think Jenny is *interested* in many things” was changed into “I think Jenny *likes* many things”). For none of the remaining items was a violation of the monotonicity assumption indicated. The items had weak to moderate scalability coefficients and a moderate scale *H* (see Table 3 for a summary).

In addition, DIF was examined for the variable “correspondence of gender and vignette.” We examined if DIF existed among students with a gender corresponding and non-corresponding vignette. This means we had four groups: girls corresponding/non-corresponding and boys corresponding/non-corresponding. The results of the analysis showed that, for items belonging to the cognitive component, no DIF appeared to occur. For the affective component, it was found that DIF occurs for both variables “gender” and “gender corresponding vignette.” For the latter DIF variable, we found that it is more likely that girls and boys with a gender-corresponding vignette show positive answers, compared with girls and boys with a non-corresponding vignette. Those results suggest the importance of a gender-corresponding vignette to overcome DIF.

First stage analyses: evaluation of the separability of the attitude components

To examine the separability of the attitude components as subscales in the three separate questionnaires, we used the AISP. The procedure was applied to all remaining items, including the adapted ones. Different boundary levels were used, ranging from  $c=0.30$  to  $c=0.60$ . The results of the final outcomes (lower bound  $c=0.30$ ) are summarized in Table 4.

The AISP was applied to all items that were included in the final scale at different boundary levels. For teachers’ and parents’ questionnaires, the results indicated a unidimensional item set as no specific subscales were distinguished. Using a high boundary level,  $c=0.60$ , the first selected scale contained a mix of items of the three components in both questionnaires.

**Table 4** Results of the automatic item step procedure of the pilot study—teachers, parents, and students

Target group		Lower bound $c=0.30$				
		Items <sup>a</sup>			$H_i$	Scale $H$
		C	A	B		
Teachers	Scale 1	5	9	6	$0.31 \leq H_i \leq 0.53$	0.41
	Scale 2	–	2	–	$0.78 \leq H_i \leq 0.78$	0.78
	Scale 3	2	–	–	$0.47 \leq H_i \leq 0.47$	0.47
	Excluded	3	–	–		
Parents	Scale 1	10	8	5	$0.31 \leq H_i \leq 0.51$	0.43
	Scale 2	1	1	1	$0.65 \leq H_i \leq 0.72$	0.69
	Excluded	4	2	–		
Students	Scale 1	1	8	9	$0.31 \leq H_i \leq 0.56$	0.47
	Scale 2	3	–	–	$0.38 \leq H_i \leq 0.43$	0.39
	Scale 3	2	–	–	$0.32 \leq H_i \leq 0.32$	0.32
	Excluded	1	–	–		

<sup>a</sup> C cognitive, A affective, B behavioral

Repeating the AISP at different boundary levels with all items showed that no distinction between the three components was made. Calculating correlations between the components and a general attitude score showed that the three components were strongly related to each other (see Table 5). The results suggest that teachers' and parents' questionnaires measure a general attitude towards inclusive education, comprising cognitive, affective, and behavioral responses. However, due to the relatively small sample sizes ( $n=58$  for teachers and  $n=58$  for the parents), the results should be interpreted with some caution.

With respect to the students' questionnaire, the results indicated a multidimensional item set. The first scale selected ( $\alpha=0.60$ ) contained only items of the affective and behavioral component. Repeating the AISP at different boundary levels with all items showed that the items belonging to the affective and behavioral component are clustered together in one scale. The items of the cognitive component were distinguished as a separate scale. The correlations between the attitude components illustrated that the affective and behavioral components are strongly related to each other (see Table 5), whereas the cognitive component is less strongly related. Hence, the results indicated that the questionnaire includes two subscales: one with items of the cognitive component and one with items of the affective and behavioral component.

#### First stage analyses: conclusion and adaptations

Outcomes of the MSA revealed that various items violated the assumptions of the MHM, and these were therefore removed from the final questionnaire. Additionally, the formulation of several items with low scalability coefficients was changed. Analysis of the separability of the attitude components revealed that the single-component model appears to apply in teachers' and parents' questionnaires. This indicates that the scales measure teachers' and parents' general attitudes towards inclusive education. Analysis of the students' questionnaire revealed that the two-component model applies with a cognitive component and an affective/behavioral component.

In addition to the results of the psychometric properties and the separability of the scale, the analyses of the data provided useful information about teachers', parents', and students' experience in completing the questionnaire. Three suggestions to further improve the scales were made:

- Change the 5-point Likert scale into a 4-point one. Initially, the questionnaire included a 5-point Likert scale with the third category representing a "neutral" choice. Because too many participants (especially 10–35% students) used this response category, the 5-point Likert scale was transformed into a 4-point scale. This invites the participants to indicate their degree of agreement in a more outspoken way.

**Table 5** Correlations of between total score and attitude components—main study

	Teacher				Parents				Students			
	T	C	A	B	T	C	A	B	T	C	A	B
Total	–				–				–			
Cognitive	0.83	–			0.93	–			0.68	–		
Affective	0.92	0.61	–		0.92	0.73	–		0.95	0.50	–	
Behavioral	0.83	0.51	0.76	–	0.85	0.66	0.82	–	0.94	0.47	0.88	–

- Add vignettes to the teacher and parent questionnaires. During the assessment of their attitudes, teachers and parents indicated having difficulty in answering the items because the general term “disability” was used. Many participants preferred a specification of the type of disability; hence, we decided to include a vignette in the teacher and parent questionnaires such as used in students’ questionnaire (see Appendix 1).
- Develop gender-specific vignettes for student questionnaires. Based on the DIF assessment, some items in the questionnaire appeared to be sensitive to the gender of the participant. During the assessment of students’ attitudes, this finding was confirmed, as some boys and girls indicated they would respond differently to an item like “I would invite Mark to my birthday party” if the child presented in the vignette would correspond with their gender. Taking the outcomes of DIF and the students’ together, we decided to develop gender-specific vignettes.

### Second stage analyses: procedure and participants

Based on the outcomes of the first stage analyses, each questionnaire was adapted and subsequently tested in a new sample. For this second stage analyses, we drew samples that were part of a wider study. For other purposes of this study, we formulated an inclusion criterion (i.e., at least one student with special needs was in grade 5–8). Taking this into account, along with experience in previous research in The Netherlands, we expected a non-response of at least 75% (Koster et al. 2009) and decided to select a large sample of regular primary schools in the north of The Netherlands ( $N=300$ ). All teachers, parents, and students of grades 5–8 of the 300 schools were invited to participate in the study by written invitation, which was mailed or emailed in February 2010. The invitation letter included a standardized return form, whereby directors and teachers were asked to indicate if they wanted to participate in the study. A total number of 26 schools wanted to participate and also met our selection criterion. The demographics of the participated teachers and parents are presented in Table 2. Prior to assessing the student attitude questionnaire, parents were invited to give their written consent to their child taking part. One parent refused permission so the child did not participate.

*Teachers* The sample study included 49 regular primary school teachers teaching in grades 5, 6, 7, or 8. The teacher of each class was asked to participate in the study and complete a questionnaire package, which could be returned by mail. Four teachers did not return the questionnaire, resulting in a sample of 45 teachers (response rate 92%).

*Students* A total number of 49 classes were part of the study, resulting in data for 938 students (51% girls). The questionnaires were personally distributed to the students by the first author and graduate students. Students were in grades 5, 6, 7, or 8 and had a mean age of 9.9 years ( $SD=0.11$ ; range, 8–12). The majority of students attended grade 6 (38%), 24% attended grade 5, and 30% and 8% of students were in grades 7 and 8, respectively.

*Parents* The parents of the participating students were also invited to participate. Each student received an envelope for their parents, which contained an invitation letter explaining the study, an invitation to participate, a questionnaire, and a return envelope. A total number of 508 parents returned the questionnaire, reflecting a response rate of 45.6%. Due to missing values, the data of 420 parents were used to evaluate the psychometric properties of the attitude questionnaire.

## Second stage analyses: evaluation of the psychometric properties

*Teacher questionnaire* The assumptions of the MHM were checked using the scalability coefficients  $H_i$  and the criteria values. A summary of the results is given in Table 6. Eight items had insufficient discrimination power and were removed from the scale (e.g., “The best way to implement inclusive education, is just by doing it”). Two items had weak scalability coefficients, but since we considered the content of the items important to cover, we decided to maintain them in the scale. DIF appeared to be absent, and none of the criteria values indicated any violation of the monotonicity assumption. The final scale included 19 items, of which six measured teachers’ beliefs, seven measured feelings, and six measured teachers’ behavioral intentions.

*Parent questionnaire* The assumptions of the MHM were checked using the scalability coefficients  $H_i$  and the criteria values. Based on the outcomes of the analysis, it was decided to delete eight items from the questionnaire (e.g., “I feel sorry for children like Alex”). One item with weak a scalability coefficient was considered to be important and therefore this item was maintained in the scale (“Regular primary schoolteachers cannot address the individual needs of children like Alex’ adequately”). Differential item functioning appeared to be absent for gender, having a child with SEN, and type of vignette. The final scale includes 24 items, of which 13 items measured parents’ beliefs, seven measured feelings, and four measured parents’ behavioral intentions (see Table 6 for a summary).

*Student questionnaire* The assumptions of the MHM were checked using the scalability coefficients  $H_i$  and the criteria values. Remarkably, all items measuring beliefs of students had insufficient scalability coefficients ( $H_i < .25$ ) and were excluded from the analysis. Taking the results of the pilot study and the main study together, we decided to remove all items of the cognitive component from the scale. This resulted in a final scale including only items of the affective and behavioral component. Differential item functioning appeared to be absent. The final scale consisted of 14 items, of which six and eight items measure students’ feelings and behavioral intentions respectively (see Table 6 for a summary).

## Second stage analyses: evaluation of the separability of the attitude components

Applying the AISP to the data of teachers’ and parents attitudes confirmed the findings of the pilot study: No distinction between the three attitude components could be made in the teacher and parent questionnaires. Using different boundary levels (ranging from  $c=0.30$  to  $c=0.60$ ) revealed a clustering of the three components into one scale. These results indicate that the questionnaire primarily measures teachers’ and parents’ general attitude towards inclusive education, which comprises cognitive, affective, and behavioral responses.

**Table 6** Summary of the final scale per target group

Target group	Deleted, $H_i < 0.30^a$			Final scale <sup>a</sup>			$H_i$	Scale $H$	Rho
	C	A	B	C	A	B			
Teachers ( $N=45$ )	4	4	0	6	7	6	$0.34 \leq H_i \leq 0.60$	0.48	0.91
Parents ( $N=420$ )	2	4	2	13	7	4	$0.15 \leq H_i \leq 0.52$	0.40	0.92
Students ( $N=938$ )	7	2	1	0	6	8	$0.37 \leq H_i \leq 0.57$	0.50	0.92

<sup>a</sup> C cognitive, A affective, B behavioral

Removing the items belonging to the cognitive component in the student questionnaire resulted in a questionnaire reflecting students' feelings and behavioral intentions. We applied to AISP to check whether the two components could be distinguished as separate subscales. Using different boundary levels revealed that the affective and behavioral items were clustered into one scale. These results indicate that the student questionnaire primarily measures general attitudes, comprising affective and behavioral responses.

## Discussion

The study presented here departed from the need for a set of good scales in order to be able to measure the attitude towards inclusive education of teachers, parents, and students in The Netherlands. A literature review showed that numerous scales to measure attitudes towards inclusive education exist, but these were all made for use in English spoken settings. The review further showed that not all scales had a sound psychometric profile and that only a minority clearly based their item composition on a one-, two-, or three-component model in attitude research. From the start, it was clear that simply selecting a scale and translating it into Dutch was no option. We used the available scales as "donor" scales. The most promising items or subscales were selected into new teacher, parent, and student scales, and these were translated and adapted to the Dutch educational situation. In practice, this resulted in making completely new questionnaires. These were constructed in two stages. The first stage analysis was focused on further improving the scales, while the second stage addressed the psychometric quality of the final scale. Data in the two stages were collected from different samples.

The results of the first stage analyses made clear that next to a number of smaller adaptations, the often used 5-point Likert scale was not optimal, that both the teacher and parent scales would benefit from adding vignettes, and that the scale for students needed gender-specific vignettes. These findings implicitly say something about the quality of many of the existing scales. Already in reviewing the available attitude scales for inclusive education, there was serious criticism about the majority of the available scales, while the findings regarding the 5-point Likert scales, the vignettes for all target groups and the gender-specific vignettes, point at further flaws in many of the existing attitude questionnaires.

The MSA applied in the first stage resulted in detailed information about the items' scalability and dimensionality structure of each questionnaire. Based on those outcomes, some adaptations were made resulting in a final questionnaire for teachers, parents, and students with appropriate discrimination power and high reliability coefficients. However, the second stage analyses still gave rise to a number of adaptations. The results furthermore supported a unidimensional item set in the teachers' and parents' questionnaires, rather than the three-component model (Triandis 1971), which is often used in attitude research. With respect to the separability of the attitude components in the students' questionnaire, we found ambiguous outcomes. Initially a two-component model with a distinction between the cognitive and affective/behavioral components was found, but repeating the procedure in the second stage analyses revealed insufficient scalability coefficients of all cognitive items. Subsequently, these items were removed, which resulted in a questionnaire including a one-component model comprising the items measuring students' feelings and behavioral intentions. This finding is in line with findings of Rosenbaum et al. (1986), who suggested that a two-component model might be more appropriate. The consequences of this finding are not quite clear yet. It is argued that beliefs and knowledge are not that important to consider, as beliefs of young students are, as yet, not stable. However, from developmental theories, it is known that children develop beliefs when they are about 4 years old (Keenan 2002), which argues for further development work on this aspect of the student scale.

The unsuccessful attempts in this study to differentiate between the three attitude components may suggest that the three-component model is a theoretical model without empirical basis. Different angles of incidence are possible here. First, it is possible that there is no such thing as a three component model and that all theoretical distinctions within the concept “attitude” inter-correlate highly with each other. A second option is that the items in the questionnaire are simply not good enough to measure the three components and that finding a one-component model is nothing more than an artifact of this weakness. Eagly and Chaiken (1993) state that “the tripartite distinction provides an important conceptual framework, that allows psychologists to express the fact that evaluation can manifest through responses of all three types, regardless of whether the types prove separable in appropriate statistical analyses” (p. 14).

Despite all these comments, this study has resulted in three new scales measuring the attitude towards inclusive education for teachers, parents, and children in The Netherlands. The past decades have made clear that implementing inclusive education is far from easy. The slow progress has been explained by, among others, pointing at teachers’, parents’, and students’ attitudes. It was stated that these were most likely rather negative. A first analysis on the data from the second stage sample does not support this: teachers, parents, and peer attitudes are overall neutral to positive. This can be interpreted as good news, but it also raises issues around socially desirable answering behavior, about placing cut-off points to decide what is negative and what is positive and about the content of the vignettes.

The issues described above point at certain limitations of the study, which need to be mentioned here. We used different procedures when drawing samples for the first and second stage, which resulted in differences in participants’ experience with children with SEN. For example, only half the teachers who participated in the first stage sample had experience with teaching students with SEN, whereas all teachers of the second stage sample had experience. As shown in other attitude studies (Avramidis and Kalyva 2007; Balboni and Pedrabissi 2000; Siperstein et al. 2007), experience with children with SEN is found to be a variable, which positively influences people’s attitudes. It is possible that the responses of the participants of the second stage sample were biased. The selection criteria we used in the second stage to select schools resulted in a small number of participating schools. Due to this specific criteria (i.e., at least one student with special needs in grades 5–8), it is likely that a large number of schools could not meet this. Moreover, it is likely that the school that did participate holds more positive attitudes than those that did not want to participate. After all, they included a student with special needs in their school/class. Another limitation relates to the distinction of the three components of student attitudes. When using a student questionnaire for other purposes, one should bear in mind that the cognitive component cannot be measured by this questionnaire.

The ultimate purpose in using an attitude questionnaire concerns the prediction of future behavior. Few studies use attitudes to predict or explain teacher and student behavior. In the study of Van der Veen et al. (2010), teacher attitudes were used to predict referral of students with special needs to special education. The outcomes of their study are in line with the meta-analysis of Glasman and Albarracín (2006), showing that attitudes can predict people’s future behavior. A measurement with good psychometric properties is necessary in such research. The attitude questionnaires developed in this current study can be used as tools for studies to predict future behavior of teachers, parents, and students in the context of inclusive education.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## Appendix 1: Vignettes used in the student questionnaire

Mark: a boy showing characteristics of AD/HD

Mark is a boy of your age and has just moved to your town. He then attends the same school as you. Mark is very noisy, he has difficulty staying in his place and walks about the classroom a lot. He also has difficulty in listening, calls out and often speaks out of turn. He has trouble working together with other children and wants to do everything his way. Mark likes playing football and he's quite good at it. Also, when Mark is angry he starts to shout, throws stuff and often leaves the classroom.

Jenny: a girl showing characteristics of a cognitive disability

Jenny is a girl of your age and has just moved to your town. She then attends the same school as you. Jenny has just started to read and write but has difficulty with maths. She can play and run like other children, but sometimes forgets the rules of certain games.. She needs extra time to learn at school and is forgetful in class. Sometimes it is difficult to understand what Jenny says. She enjoys playing music. For part of the day, Jenny receives extra learning assistance outside the classroom.

John: a boy showing characteristics of a physical disability

Mark is a boy of your age and has just moved to your town. He then attends the same school as you. John has difficulty walking. He walks with braces around his legs, uses crutches and sometimes needs a wheelchair for daytrips with his family. John often needs to skip school as he visits a doctor who helps him with his walking. John is a good learner and he's funny. He uses a computer at school because he's a slow writer. Sometimes it is difficult to understand what he says.

## References

- Ajzen, I. (2005). *Attitudes, personality and behavior*. New York: Open University Press.
- Albarracín, D., Johnson, B. T., & Zanna, M. P. (2005). *The handbook of attitudes*. Mahwah: Erlbaum.
- Alderfer, M. A., Wiebe, D. J., & Hartmann, D. P. (2001). Social behaviour and illness information interact to influence the peer acceptance of children with chronic illness. *British Journal of Health Psychology*, *6*, 243–255.
- Avramidis, E., Bayliss, P., & Burden, R. (2000). A survey into mainstream teachers' attitudes towards the inclusion of children with special educational needs in the ordinary school in one local education authority. *Educational Psychology*, *20*(2), 191–211.
- Avramidis, E., & Kalyva, E. (2007). The influence of teaching experience and professional development on Greek teachers' attitudes towards inclusion. *European Journal of Special Needs Education*, *22*(4), 367–389.
- Bagozzi, R. P., & Burnkrant, R. E. (1979). Attitude organization and the attitude-behavior relationship. *Journal of Personality and Social Psychology*, *37*, 913–929.
- Bagozzi, R. P., & Burnkrant, R. E. (1985). Attitude organization and the attitude-behavior relation: A reply to Dillon and Kumar. *Journal of Personality and Social Psychology*, *49*(1), 47–57.
- Balboni, G., & Pedrabissi, L. (2000). Attitudes of Italian teachers and parents toward school inclusion of students with mental retardation: The role of experience. *Education and Training in Mental Retardation and Developmental Disabilities*, *35*, 148–159.
- Beck, A. R., Bock, S., Thompson, J. R., Bowman, L., & Robbins, S. (2006). Is awesome really awesome? How the inclusion of informal terms on an AAC device influences children's attitudes toward peers who use AAC. *Research in Developmental Disabilities*, *27*(1), 56–69.
- Breckler, S. J. (1984). Empirical validation of affect, behavior, and cognition as distinct components of attitude. *Journal of Personality and Social Psychology*, *47*, 1191–1205.
- Chow, P., & Winzer, M. M. (1992). Reliability and validity of a scale measuring attitudes toward mainstreaming. *Educational and Psychological Measurement*, *52*(1), 223.
- De Boer, A., Pijl, S. J., & Minnaert, A. (2010). Attitudes of parents towards inclusive education: A review of the literature. *European Journal of Special Needs Education*, *25*(2), 165–181.



- De Boer, A. A., Pijl, S. J., & Minnaert, A. E. M. G. (2011). Regular primary school teachers' attitudes towards inclusive education: A review of the literature. *International Journal of Inclusive Education*, 15(3), 331–353.
- Dillon, W.R. & Kumar, A. (1985). Attitude organization and the attitude-behavior relation: A critique of Bagozzi and Burnkrant's reanalysis of Fishbein and Ajzen. *Journal of Personality and Social Psychology*, 49(1), 33–46.
- Drenth, P. J. D., & Sijtsma, K. (2006). Betrouwbaarheid. In *Testtheorie: Inleiding in de theorie van de psychologische test en zijn toepassingen* (pp. 190–253). Houten: Bohn Stafleu van Loghum.
- Eagly, A. H., & Chaiken, S. (1993). The nature of attitudes. In A. H. Eagly & S. Chaiken (Eds.), *The psychology of attitudes* (pp. 1–21). Fort Worth: Harcourt Brace College.
- Ferguson, D. L. (2008). International trends in inclusive education: The continuing challenge to teach each one and everyone. *European Journal of Special Needs Education*, 23(2), 109–120.
- Fishbein, M., & Ajzen, I. (1974). Attitudes towards objects as predictors of single and multiple behavioral criteria. Erratum. *Psychological Review*, 81(2), 164.
- Flem, A., & Keller, C. (2000). Inclusion in Norway: A study of ideology in practice. *European Journal of Special Needs Education*, 15(2), 188–205.
- Glasman, L. R., & Albarracín, D. (2006). Forming attitudes that predict future behavior: A meta-analysis of the attitude-behavior relation. *Psychological Bulletin*, 132(5), 778–822.
- Keenan, T. (2002). *An introduction to child development*. London: Sage.
- Koster, M., Timmerman, M. E., Nakken, H., Pijl, S. J., & Van Houten, E. J. (2009). *European Journal of Psychological Assessment*, 25(4), 213–222.
- Mahat, M. (2008). The development of a psychometrically-sound instrument to measure teachers' multidimensional attitudes toward inclusive education. *International Journal of Special Education*, 23(1), 82–92.
- Meijer, C. J. W. (2003). *Special needs education across Europe*. Middelfart: European Agency for Development in Special Needs Education.
- Mokken, R. J. (1971). *A theory and procedure of scale analysis*. The Hague/Berlin: De Gruyter Mouton.
- Molenaar, I. W., & Sijtsma, K. (2000). *User's manual MSP5 for Windows*. Groningen: iecProGAMMA.
- Norwich, B. (1994). The relationship between attitudes to the integration of children with special educational needs and wider socio-political views: A US-English comparison. *European Journal of Special Needs Education*, 9, 91–106.
- Nowicki, E. A. (2006). A cross-sectional multivariate analysis of children's attitudes towards disabilities. *Journal of Intellectual Disability Research*, 50(5), 335–348.
- Ostrom, T. M. (1969). The relationships between the affective, behavioural and cognitive components of attitude. *Journal of Experimental Social Psychology*, 5, 12–30.
- Rosenbaum, P. L., Armstrong, R. W., & King, S. M. (1987). Parental attitudes toward children with handicaps: New perspectives with a new measure. *Journal of Developmental and Behavioral Pediatrics*, 8(6), 327–334.
- Rosenbaum, P. L., Armstrong, R. W., & King, S. M. (1986). Children's attitudes toward disabled peers: A self-report measure. *Journal of Pediatric Psychology*, 11, 517–530.
- Scheepstra, A. J. M., Nakken, H., & Pijl, S. J. (1999). Contacts with classmates: The social position of pupils with Down's syndrome in Dutch mainstream education. *European Journal of Special Needs Education*, 14(3), 212–220.
- Siperstein, G. N., Parker, R. C., Bardón, J. N., & Widaman, K. F. (2007). A national study of youth attitudes toward the inclusion of students with intellectual disabilities. *Exceptional Children*, 73, 435–455.
- Stoiber, K. C., Gettinger, M., & Goetz, D. (1998). Exploring factors influencing parents' and early childhood practitioners' beliefs about inclusion. *Early Childhood Research Quarterly*, 13(1), 107.
- Sijtsma, K., & Molenaar, I. W. (2002). *Introduction to nonparametric item response theory* (Vol. 5). London: Sage Publications.
- Tafa, E., & Manolitsis, G. (2003). Attitudes of Greek parents of typically developing kindergarten children towards inclusive education. *European Journal of Special Needs Education*, 18(2), 155.
- Triandis, H. C. (1971). Attitudes measurement and methodology. In *Attitudes and attitude change* (pp. 26–59). New York: Wiley.
- Van der Veen, Smeets, E., & Derricks, M. (2010). Children with special educational needs in The Netherlands: Number, characteristics and school career. *Educational Research*, 52(1), 15–43.
- Vignes, C., Coley, N., Grandjean, H., Godeau, E., & Arnaud, C. (2008). Measuring children's attitudes towards peers with disabilities: A review of instruments. *Developmental Medicine and Child Neurology*, 50(3), 182–189.
- Vislie, L. (2003). From integration to inclusion: Focusing global trends and changes in the western European societies. *European Journal of Special Needs Education*, 18(1), 17.