

---

## Interpersonal Teacher Behaviour and Student Outcomes

Perry den Brok<sup>1</sup>, Mieke Brekelmans<sup>1</sup>, and Theo Wubbels<sup>2</sup>

<sup>1</sup>IVLOS Institute of Education, Utrecht University, The Netherlands,

and <sup>2</sup>Institute of Educational Sciences, Utrecht University, The Netherlands

---

### ABSTRACT

In this study, the effectiveness of secondary education teachers' interpersonal behaviour is investigated by analysing data from 2 samples: a study on 45 Physics teachers and their 3rd-year classes and a study on 32 English as a Foreign Language (EFL) teachers and their 3rd-year classes. Teacher interpersonal behaviour was studied by means of students' perceptions of this behaviour, collected with the Questionnaire on Teacher Interaction (QTI). These perceptions include 2 important dimensions: Influence and Proximity. Results of multilevel analyses with various covariates indicated that Influence and Proximity were positively related to both cognitive and affective outcomes. Interpersonal behaviour explained up to more than half of the variance in student outcomes at the teacher-class level. The outcomes suggest that interpersonal behaviour as perceived by students may be an important variable for educational effectiveness researchers.

---

### INTRODUCTION

Since its start, the domain of educational effectiveness research has studied the link between teaching and the outcomes of students. This interest in the effects of teacher behaviour resulted in the development of a new subdomain: *instructional effectiveness* (Creemers & Reezigt, 1996). Research on educational and instructional effectiveness has shown that between 7 and 15% of the variance in student outcomes is related to differences between schools, teachers, and classes. Most of this percentage is due to differences between

---

Address correspondence to: P.J. den Brok, IVLOS Institute of Education, Utrecht University, P.O. Box 80127, 3508 TC Utrecht, The Netherlands. Tel.: +31-30-2532231. Fax: +31-30-2532741. E-mail: p.denbrok@ivlos.uu.nl

Manuscript submitted: April 28, 2003

Accepted for publication: April 19, 2004

*teachers* (Reynolds, 1995). Teacher behaviours that have been found to relate to student outcomes are, for instance, clarity, feedback, classroom management, and communication of teacher expectations (Creemers, 1994). The domain of effectiveness research has been particularly important for its statistical and methodological contributions: By employing multilevel analysis techniques and including many covariates in analyses, effects of teaching could be estimated very precisely (Scheerens & Bosker, 1997).

The present study tries to add to the current knowledge base on educational effectiveness by linking concepts, methods, and instruments from the domain of *research on teaching and learning environments research* to models and analysis techniques from the *educational effectiveness* domain.

An important characteristic of our study is the use of student ratings to measure teacher behaviour. Until recently, educational effectiveness research mainly focussed on observable behaviours of teachers (Lowyck, 1994). During the last decade, researchers have started to use teacher and student questionnaires to measure *teachers'* and *students' perceptions* of teacher behaviour. In at least six publications that appeared in the journal of *School Effectiveness and School Improvement* between 1995 and 2002 that included teacher behaviour in their design, student or teacher questionnaires were used (e.g., d'Agostino, 2000; Kyriakides, Gagatsis, & Campbell, 2000; Luyten & De Jong, 1998; Sandal, Wold, & Bronis, 1999; Van Damme, De Fraine, Van Landeghem, Opdenakker, & Ongenha, 2002; Van de Grift, Houtveen, & Vermeulen, 1997). Several arguments can be provided for the (additional) use of student perceptions. First, psychologically oriented research has shown that the effect teachers have on students is determined by the students' psychological response to what the teacher does (Doyle, 1979; Shuell, 1996; Shulman, 1986), rather than by anything else. According to Shuell (1996), the way in which learners perceive, interpret, and process information in the instructional situation (including content and social processes) is crucial in determining what the student will learn. Second, student perceptions are cheap and efficient to gather. A third argument is that the experience of students with the behaviour of a certain teacher is often based on a large amount of lessons (Den Brok, 2001; Fraser, 1998). Therefore, student perceptions can account for the history characteristic of the classroom context (e.g., Doyle, 1986; Shuell, 1996). Fourth, student perceptions often consist of the composite judgement of all the students in a class. Therefore, student perceptions that have been averaged over a class are only marginally subject to mood swings, personal preferences, and other personal or situational factors

(Den Brok, 2001). Last, students have an advantage in judging classroom environments because they have encountered many different situations and contexts, which may help to describe a differentiated picture (Wubbels & Brekelmans, 1998). In this article, perception data are analysed from 3rd-year students in secondary education (14–15 years of age). It has been shown that secondary education students, as well as students at the end of primary education, are able to provide ratings of teacher behaviour that are sufficiently stable, reliable, valid, and predictive for teacher evaluation and research purposes (Driscoll, Peterson, Crow, & Larson, 1985; Mak, 2001; Peterson & Stevens, 1988; Scriven, 1994; Taba, Tylor, & Smith, 1998).

The domain of classroom environment research (e.g., Fraser, 1994, 1998), which found its origin in early instructional effectiveness studies and studies on the interaction between person and environment (Moos, 1979; Murray, 1938; Walberg, 1979), has largely used student perceptions to study teaching. From this domain, much has been learned about the variety of teaching in terms of student and teacher perceptions (Levy, Den Brok, Wubbels, & Brekelmans, 2003). Within this specific field, researchers have also been able to find consistent relationships between teaching and student outcomes (Brekelmans, Wubbels, & Den Brok, 2002; Den Brok, 2001). However, learning environments research has been criticised for the methodology used to study teaching: It hardly included covariates or employed multilevel analysis (Fraser, 1998), resulting in overestimation of teacher effects. It is believed that by combining insight from both domains, research on the relationship between teacher behaviour and student outcomes may take new directions and that additional arguments may be found to strengthen discussions on the importance of teaching within each particular domain.

Another specific feature of our study is the theoretical framework used to conceptualise teaching. We believe that individuals in the classroom environment and what they learn are influenced by a variety of (interpersonal, emotional, cultural) factors in addition to the cognitive factors associated with classroom learning (e.g., Shuell, 1996). In this context, the teacher is one of the elements contributing to the opportunities for pupils to learn. Teaching can be studied in terms of various different perspectives, such as a *subject-content perspective* that analyses teaching from the specific situation of the subject matter, a *learning activities perspective* that describes teaching in terms of the way the teacher elicits learning activities with pupils, an *interpersonal perspective* that describes teaching in terms of the relationship between

teacher and pupils, a *moral perspective* describing teaching in terms of the values a teacher is communicating to pupils, and an *organisational perspective* focusing on the teacher as a member of the school organisation (e.g., Brekelmans, Slegers, & Fraser, 2000). In this study, we focus on the interpersonal perspective on teaching, which means that *teacher behaviour* is described and measured in terms of the teacher-student relationship. As such, our study adds to the existing knowledge base, because it investigates associations between learning outcomes and classroom management behaviours from a *relational* viewpoint (e.g., Van Damme et al., 2002), in addition to *methodical issues or behaviours*, such as grouping procedures and classroom organisation (e.g., Brophy & Good, 1986; Creemers, 1994; Lee, 1995).

In the present study, the effectiveness of secondary school teachers' behaviour is investigated by analysing data from two separate research efforts: a study on 45 Physics teachers and their 3rd-year classes (Brekelmans, Wubbels, & Créton, 1990) and a study on 32 English as a Foreign Language (EFL) teachers and their 3rd-year classes (Den Brok, 2001). Both datasets used include cognitive as well as affective (subject-specific motivation) student outcomes. The Physics teachers' dataset was gathered some time ago, but that probably does not mean that it is outdated: A more recent study with a similar sample showed that their interpersonal behaviour had hardly changed over the past 10 years (Wubbels & Brekelmans, 1997). This may suggest that any conclusions drawn from the Physics dataset are probably still valid.

By investigating the relationship between teacher behaviour and two different outcome measures for two different datasets, the study also adds to two other discussions within the domain of effectiveness research. First, in the past, effectiveness research was criticised for its exclusive focus on cognitive learning outcomes (Creemers & Scheerens, 1994; Scheerens, 1993). Recent studies have employed multiple student outcomes, as does the present study by including both cognitive and affective learning outcomes. Second, the number of effectiveness studies investigating the subject of (modern) foreign languages is rather small. Some of these studies have indicated lower amounts of variance in *student outcomes* for the languages than for other subjects – like Mathematics or Physics – and, subsequently, lower amounts of variance to be explained at the class or teacher level (Grisay, 1996; Hill & Rowe, 1996; Luyten, 1994; Mandeville & Anderson, 1987; Scheerens & Creemers, 1996; Willms & Raudenbush, 1989). Luyten (1998) reviewed studies on stability of effectiveness across subjects in primary and secondary education. Outcomes of his review indicate that not many studies have included language teaching,

either first or second language (a similar conclusion is drawn by Scheerens & Creemers, 1996, for the Netherlands). By including datasets from two subjects, one of which is a second language, the study tries to expand the discussion on teacher effects across different subjects.

The next section starts with a description of the model used to study teacher behaviour from an interpersonal perspective. Next, research is discussed that has investigated the relationship between teacher behaviour – in terms of this model – and student outcomes. This discussion includes strengths and weaknesses of these investigations. After a presentation of the research questions, outcomes of multilevel analyses on two datasets are discussed, one dataset of EFL teachers and one dataset of Physics teachers.

### INTERPERSONAL TEACHER BEHAVIOUR

In our conceptualisation of the interpersonal perspective on teaching, some concepts of the so-called systems approach to communication (Watzlawick, Beavin, & Jackson, 1967) are important. In line with the systems approach to communication, we conceive classroom groups as ongoing systems. For ongoing systems a certain stability is important for their continued existence. When students meet a teacher in a new class, they will be relatively open to any impression the teacher can make. Relatively, because the context of the classroom will raise certain (stereotypical) expectations for the role of the teacher. After the first lesson, the students will have tentative ideas about the pattern of relationship with this particular teacher, based on experiences during the first lesson. The second lesson the teacher may behave differently and students may consequently adjust their ideas about the teacher. After a few lessons in a new class, tentative ideas about the teacher will have stabilised and students can tell what kind of teacher someone “is”. This stability of perceptions equally applies to the teacher’s ideas about the students. Once the tone is set, it is difficult to modify. Both students and teachers resist against changes (see also Blumenfeld & Meece, 1985; Doyle, 1986). To describe these kinds of processes, the systems approach to communication distinguishes different levels of communication. The lowest level consists of messages, one question, assignment, response, gesture, et cetera. The intermediate level is that of interactions, chains of several messages. When the interactions show recurrent patterns and some form of regularity, one has arrived at the pattern level. It is this pattern level that is

important in describing the rather stable interpersonal relationships that determine the working atmosphere of classrooms. We will focus on this last level.

In the systems approach to communication, the focus is on the effect of communication on the persons involved (pragmatic aspect). This pragmatic orientation is characterised in our conceptualisation of the interpersonal perspective by means of focus on the perception of students of their teacher's behaviour.

To be able to describe the perceptions students have of the behaviour of their teacher, Wubbels, Créton, and Hooymayers (1985; see Wubbels & Levy, 1993) developed a model. They applied a general model for interpersonal relationships designed by Leary (1957) to the context of education. The Leary model has been extensively investigated in clinical psychology and psychotherapeutic settings (Strack, 1996). It has proven to be a rather complete model to describe interpersonal relationships (see e.g., Foa, 1961; Lonner, 1980). In the Leary model, two dimensions are important. Leary called them the Dominance-Submission axis and the Hostility-Affection axis. Whereas the two dimensions have occasionally been given other names – Brown (1965) used Status and Solidarity, Dunkin and Biddle (1974) Warmth and Directivity – they have generally been accepted as universal descriptors of human interaction. The two dimensions have also been easily transferred to education. Slater (1962) used them to describe pedagogical relationships, and Dunkin and Biddle (1974) demonstrated their importance in teachers' efforts to influence classroom events.

Adapting the Leary Model to the context of education, Wubbels et al. (1985) used the two dimensions, which they called *Influence* (Dominance-Submission) and *Proximity* (Opposition-Cooperation) to structure the perception of eight behaviour segments: leadership, helpful/friendly behaviour, understanding behaviour, giving students freedom, uncertain, dissatisfied, admonishing, and strict behaviour. Figure 1 is a graphic representation of the model of Wubbels et al. (1985), the *Model for Interpersonal Teacher Behaviour*.

The Model for Interpersonal Teacher Behaviour (see Fig. 1), as well as the Leary model, are special models because of their statistical properties, and are theoretically linked to a particular branch of models called *circumplex models* (e.g., Blackburn & Renwick, 1996; Fabrigar, Visser, & Browne, 1997; Gaines et al., 1997; Gurtman & Pincus, 2000). Circumplex models assume that the eight interpersonal sectors can be represented by two, independent dimensions (Influence and Proximity), are ordered with

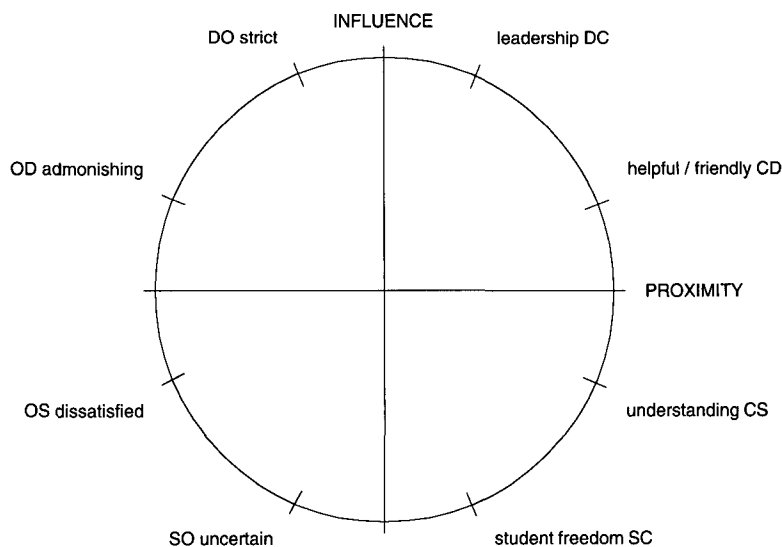


Fig. 1. The Model for Interpersonal Teacher Behaviour.

equal distances to each other on a circular structure, and maintain equal distances to the middle of the circle.

The sections are labelled DC, CD, et cetera, according to their position in the co-ordinate system. For example, the two sectors *leadership* and *helpful/friendly* are both characterised by Dominance and Cooperation. In the DC sector, the Dominance aspect prevails over the Cooperation aspect. A teacher displaying DC behaviour might be seen by students as enthusiastic, a good leader, and the like. The adjacent CD sector includes behaviours of a more cooperative and less dominant type; the teacher might be seen as helpful, friendly, considerate.

#### REVIEW OF RESEARCH ON INTERPERSONAL BEHAVIOUR AND STUDENT OUTCOMES

Classroom environment and effectiveness studies that have included interpersonal teacher behaviour usually indicate a strong and positive association between perceptions of Influence and Proximity or their related (sub)sectors and *cognitive student outcomes*. In the study by Brekelmans et al. (1990) for

example, student perceptions of teacher Influence were related to cognitive outcomes. The higher a teacher was perceived on the Influence dimension, the higher the outcomes of students were on a Physics test. In their study, teacher Influence was the most important variable at the class level; other variables included at the class level were curricular material, the number of lessons per week, class size, type of education, and percentage of boys in the class. Other studies found positive correlations or regression coefficients for the leadership sector and cognitive student outcomes (Goh, 1994; Henderson, 1995). However, the effects found in these studies were probably overestimated, because they were not corrected for the influence of other covariates and because the nested structure of the data in these studies was not taken into account.

Similar relationships have also been found for the Proximity dimension and Proximity related sectors such as helpful/friendly and understanding, and to a lesser degree student responsibility/freedom (Evans, 1998; Goh, 1994; Henderson, 1995). The more teachers were perceived as co-operative, the higher students' scores on cognitive tests. As was the case with the Influence dimension, the effects of Proximity in these studies may have been overestimated due to methods of analyses used and the absence of covariates. However, relationships between Proximity and cognitive outcomes are not always straightforward. Sometimes it can only be proven that opposition, or uncertainty, dissatisfaction, and admonishing behaviour lead to lower performance, but not that leadership, friendliness, and understanding behaviour lead to higher performance (Rawnsley, 1997). In still other cases, the relationship between Proximity/Immediacy and cognitive outcomes is not linear but curvilinear (i.e., lower perceptions of proximity lead to lower outcomes, but intermediate and higher values lead to higher performance until a certain ceiling of optimal Proximity has been reached). If report card grades have been used as outcome measures, relationships with interpersonal behaviour are inconclusive (Levy, Wubbels, & Brekelmans, 1992; Van Amelsvoort, 1999). No relationship between student perceptions of teacher Proximity and Influence and their report card grades was found in these studies.

Studies investigating associations between the teacher-student relationship and *affective outcomes*, display a much more consistent pattern than studies investigating the relationship with cognitive outcomes. All studies find a positive association of both Influence and Proximity with affective outcome



measures, usually measured in terms of subject-specific motivation. Generally, effects of Proximity are somewhat stronger than effects of Influence. In a study with Physics teachers and their students, Brekelmans et al. (1990) found a clear relationship between Proximity and student motivation for Physics. The higher the perception of Proximity, the higher the motivation of the students. With more specific measures of students' subject-specific motivation, other studies found positive relationships for helpful/friendly and understanding behaviour with pleasure, confidence, effort, and relevance of students (Brekelmans et al., 2002; Van Amelsvoort, 1999).

Strong and positive associations have also been demonstrated between several interpersonal sectors, such as leadership and helpful/friendly, and affective outcomes, whereas negative relationships have been found with admonishing, dissatisfied, and, in most cases, strictness (Evans, 1998; Goh, 1994; Henderson, 1995; Rawnsley, 1997; Van Amelsvoort, 1999). The weakest associations have been found between interpersonal behaviour and confidence (Van Amelsvoort, 1999). Van Amelsvoort also demonstrated that the association of interpersonal teacher behaviour with students' subject-specific motivation is both direct as well as indirect via student motivation and regulation processes.

Only a small number of studies have investigated the associations of interpersonal teacher behaviour with student outcomes, taking into account other perspectives on teaching (e.g., other teacher behaviour variables). Most of these were done in Australia (Evans, 1998; Goh, 1994; Henderson, 1995; Rawnsley, 1997), one study was found in the Netherlands (Van Amelsvoort, 1999). Some studies have found similar amounts of variance explained by interpersonal teacher behaviour as compared to other teacher behaviours with respect to examination scores (Goh, 1994; Henderson, 1995). One study, investigating outcomes on a practical test, found a larger amount of variance explained by interpersonal teacher behaviour (Henderson, 1995), whereas another study found higher amounts of variance explained by other teaching variables (Rawnsley, 1997). The amounts of variance shared by interpersonal teacher behaviour and other teacher behaviours were rather low (less than 5%) in all of the studies mentioned. This means that interpersonal teacher behaviour has a separate, distinctive relationship with cognitive student outcomes. Studies investigating the relations of interpersonal behaviour simultaneously with other teaching variables with affective outcomes, have found similar amounts of variance explained as those investigating cognitive outcomes. Two

studies indicated similar amounts of variance explained by interpersonal and other teacher behaviours (Henderson, 1995; Rawnsley, 1997) and one study reported larger amounts of variance explained by interpersonal behaviour (Goh, 1994). All three studies reported that much of the variance was shared by all teacher behaviours, rather than explained by only interpersonal or other teacher behaviours.

A small number of studies have investigated consistency of interpersonal teacher behaviour across subjects (taught by different teachers). These found significant differences between (foreign) language teachers and teachers of other subjects. Generally, studies demonstrated higher amounts of influence and lower amounts of proximity for language teachers as compared to their colleagues of other subjects (Levy et al., 1992; Van Amelsvoort, 1999).

Research on the relationship between interpersonal teacher behaviour and student outcomes displays fairly consistent results, but the studies are subject to some limitations. First, with the exception of the Brekelmans et al. (1990) study, none of the studies employed multilevel analysis techniques. Instead, one-way analyses of variance (ANOVA), multivariate analyses of variance or correlations were used to investigate associations. These analysis techniques can provide useful information, but they usually overestimate effects found, because they assume random sampling. In most studies, classes were sampled as a whole, meaning that data were hierarchical in nature. Hox (1995, following Muthén, 1994) argues that answers to questions of respondents in a class may be more similar than those of students that are randomly sampled, since students in a class have similar experiences and history and find themselves in similar contexts. Ignoring this may lead to spurious correlations or conclusions based on artificially high associations. Second, in most cases the effects found were not corrected for covariates, such as student, teacher, or class characteristics. In some cases, corrections were only limited to (a small number of) other teacher behaviours. Again, this may have led to overestimation of effects of interpersonal teacher behaviour on student outcomes (Levy et al., 2003). Third, many of the studies operationalised teaching in terms of the eight sectors of the model for interpersonal teacher behaviour, whereas only a few used the two underlying dimensions of influence and proximity. The interpersonal dimensions are preferable from a research point of view, because they are (theoretically) independent and can be used separately (whereas the eight sectors are interrelated), and because they are less subject to reliability and validity problems (e.g., Den Brok, Fisher, et al., 2003).

## RESEARCH QUESTIONS

This study investigates the association between interpersonal teacher behaviour and EFL and Physics students' cognitive and affective outcomes.<sup>1</sup> The association is further investigated with two research questions:

1. What variance distribution between students, classes, and teachers is present in EFL and Physics students' cognitive and affective outcomes?
2. How much variance in EFL and Physics students' cognitive and affective outcomes can be explained by teacher interpersonal behaviour?

## METHOD

## Variables

*Interpersonal Teacher Behaviour*

Data about the perceptions of students on the teacher-student relationship have been gathered by means of the *Questionnaire on Teacher Interaction (QTI)*. The Dutch version of the QTI consists of 77 items which are answered on a

Table 1. Number of Items and a Typical Item for the QTI-Scales.

Scale	Number of items	Typical item
DC Leadership	10	S/he is a good leader
CD Helpful/friendly	10	S/he is someone we can depend on
CS Understanding	10	If we have something to say s/he will listen
SC Student responsibility/freedom	9	S/he gives us a lot of free time in class
SO Uncertain	9	S/he seems uncertain
OS Dissatisfied	11	S/he is suspicious
OD Admonishing	9	S/he gets angry
DO Strict	9	S/he is strict

<sup>1</sup>Results describing the association between students' perceptions of teacher interpersonal behaviour and student outcomes for the Physics data have been published elsewhere (Brekelmans et al., 1990). However, in this study we reanalysed the data to optimise comparison of results with the EFL data.

5-point Likert scale. These items are divided into eight scales which conform to the eight sectors of the model. Table 1 presents a typical item and the number of items for each scale.

Several studies have been conducted on the reliability and validity of the QTI. They have included Dutch (e.g., Brekelmans et al., 1990; Den Brok, 2001; Wubbels et al., 1985), American (Wubbels & Levy, 1991), and Australian (Fisher, Fraser, & Wubbels, 1992) samples. Both reliability and validity were satisfying.<sup>2</sup>

Each completed questionnaire yields a set of eight scale scores. Scale scores of students from the same class are combined into a class mean. In the study presented in this chapter we analyse the teacher-student relationship on the basis of dimension scores. To summarise the scale scores by means of dimension scores we use linear combinations of the scale scores.<sup>3</sup> We designate the two linear combinations of the eight scores as an Influence score and a Proximity score. The higher these scores are, the more dominance or cooperation is perceived in the behaviour of a teacher.

Reliability was computed for each of the scales of the QTI: multilevel  $\lambda$  (Snijders & Bosker, 1999) and Cronbach's  $\alpha$  for data aggregated to the class level. Cronbach's  $\alpha$  represents consistency across items, whereas multilevel  $\lambda$  represents consistency across groups of students. The results are presented in Table 2.

Reliability coefficients (both  $\lambda$  and  $\alpha$ ) are very high (around .90). The reliability of the scales Giving Responsibility/Freedom (SC) and Strict (DO) are somewhat lower, and the reliability of the scale Understanding (CS) is somewhat higher.

---

<sup>2</sup>The homogeneity of each of the eight groups of items was considerable. The internal consistencies (Cronbach's  $\alpha$ ) at class level are generally above .80. The agreement between the scores of students in a single class met the general requirements for agreement between observer scores. The mean of the internal consistencies was .92 (Cronbach's  $\alpha$ ; students' scores in one class were considered as repeated measures). Factor analyses on class means and LISREL analyses (Den Brok, 2001; Den Brok, Levy, Wubbels, & Rodriguez, 2003; Wubbels & Levy, 1991) determined that the two-factor structure did indeed support the eight scales. Brekelmans et al. (1990) demonstrated that both factors explain 80% of the variance on all the scales of the Dutch QTI. Similar results were obtained for the American version (Wubbels & Levy, 1991).

<sup>3</sup>To this end the eight scores are represented as vectors in a two-dimensional space, each dividing a section of the model of interpersonal behaviour in two and with a length corresponding to the height of the scale score. We then compute the two coordinates of the resultant of these eight vectors.

Table 2. Reliability, Multilevel  $\lambda$  and Cronbach's  $\alpha$  (Class Level) of the QTI.

Scale	Physics sample		EFL sample	
	$\lambda$	$\alpha$	$\lambda$	$\alpha$
DC	.93	.94	.95	.94
CD	.90	.96	.94	.97
CS	.88	.97	.93	.98
SC	.85	.84	.90	.91
SO	.93	.90	.94	.93
OS	.86	.91	.92	.92
OD	.89	.90	.93	.94
DO	.89	.88	.94	.92

*Note.* See Figure 1 for the meaning of the scales.

Table 3. Intraclass Correlations of the QTI-Scales.

Scale	Physics sample	EFL sample
DC	.56	.46
CD	.47	.55
CS	.42	.47
SC	.36	.41
SO	.57	.41
OS	.37	.40
OD	.45	.42
DO	.45	.41

*Note.* See Figure 1 for the meaning of the scales.

With Mplus (Muthén & Muthén, 1999) intra-class correlations of the scales were computed. The intra-class correlations, which indicate what amount of variance of the QTI is located at the between level, are listed in Table 3.

With respect to the discriminant validity, we calculated and found the percentages of variance at the between level (teacher-class level) to be between 36 and 57 for the Physics teachers' data and between 40 and 55 for the English teachers' data. These percentages are rather high compared to other instruments that measure perceptions of people or objects in clustered or interdependent situations (see also Wubbels & Levy, 1993).

Construct validity of the QTI was investigated by subjecting the scale scores to a multilevel factor analysis using Mplus. In these analyses, a model was fitted that assumed the eight sectors of the QTI to be ordered in a circle

and to be represented by two independent dimensions. Model fit statistics for EFL were: Chi-squared = 72.15 (with  $df = 13$ ;  $p = .00$ ); CFI = .99; TLI = .94; RMSEA = .06 and SRMR = .04; for Physics: Chi-squared = 50.59 (with  $df = 13$ ;  $p = .00$ ); CFI = .98; TLI = .91; RMSEA = .08 and SRMR = .04. Whereas model fit was satisfactory in terms of CFI, TLI, and SRMR, RMSEA indicates that further improvement is possible. Moreover, the outcomes suggest that sectors are not equally distributed over the circle and equally distanced to the circle center. The factor loadings resulting from this model are presented in Table 4, and they are graphically displayed in Figure 2. As can be seen, the factor loadings more or less follow a circular ordering, but they are clearly not equally spaced in the circle. To investigate the consequences of these dislocations for the two interpersonal dimensions (Influence and Proximity), we computed correlations between dimension scores based on the *theoretical* structure (as displayed in Fig. 1) and scores based on the *empirical* structure (as displayed in Fig. 2,<sup>4</sup> and using the factor loadings of Table 4). Correlations were very high for both datasets: .989 (Influence) and .998 (Proximity) for Physics; .987 (Influence) and .999 (Proximity) for EFL. These outcomes provide partial support for the Model of Interpersonal Teacher Behaviour and suggest using the two dimension scores, rather than the eight

Table 4. Factor Loadings for the Unequally Spaced Circumplex Model.

	Physics sample		EFL sample	
	Factor 1	Factor 2	Factor 1	Factor 2
DC	1.00	.35	1.00	.56
CD	.36	1.06	.25	1.15
CS	.09	1.00	.02	1.00
SC	-.37	.58	-.44	.53
SO	-1.00	.11	-1.00	-.16
OS	-.14	-.83	-.05	-.73
OD	-.04	-.86	.25	-.78
DO	.40	-.63	.56	-.58

Note. See Figure 1 for the meaning of the scales.

<sup>4</sup>Theoretical factor scores on Influence and Proximity were computed as follows:

Influence =  $(.92 * DC) + (.38 * CD) - (.38 * CS) - (.92 * SC) - (.92 * SO) - (.38 * OS) + (.38 * OD) + (.92 * DO)$ ; Proximity =  $(.38 * DC) + (.92 * CD) + (.92 * CS) + (.38 * SC) - (.38 * SO) - (.92 * OS) - (.92 * OD) - (.38 * DO)$ .

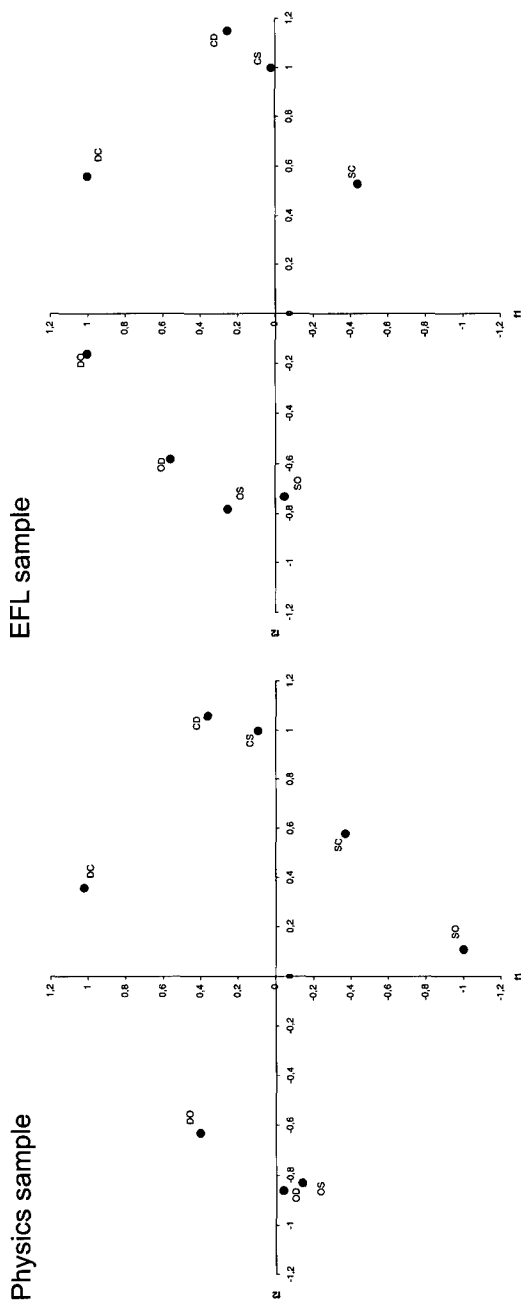


Fig. 2. Graphical presentation of the class-level factor loadings of the Physics and the EFL sample.

sector scores, in establishing the relationship between interpersonal teacher behaviour and students' outcomes.

### *Student Outcomes*

One difference between the two studies is the measurement of the outcome variables. The Physics dataset included a cognitive problem-solving test with 23 multiple choice questions, and the English sample contained a reading comprehension test with 20 items. In both cases, item scores were transformed into dichotomous variables (right/wrong categories). Reliability of both tests was sufficient: the KR-20 was .82 for EFL, and it was .65 for Physics. To obtain similar scale scores, test scores were transformed into a new variable with a range between 0 and 10. Also, slightly different instruments had been used to determine affective outcomes. Affective outcomes were measured in terms of students' subject-specific motivation (Clément, Dörnyei, & Noels, 1994; Gardner & Lambert, 1972; Gardner & MacIntyre, 1993). The instrument for Physics consisted of 30 items (e.g., Brekelmans et al., 1990) and included five aspects: pleasure ( $\alpha = .91$ ), confidence ( $\alpha = .91$ ), effort ( $\alpha = .85$ ), relevance ( $\alpha = .91$ ), and structuredness ( $\alpha = .86$ ). Because structuredness was not part of the EFL motivation instrument, it was not used for the present study. The instrument for EFL consisted of 32 items and four aspects: pleasure ( $\alpha = .87$ ), relevance ( $\alpha = .77$ ), confidence ( $\alpha = .86$ ), and effort ( $\alpha = .69$ ). Confirmatory factor analyses on both instruments (Den Brok, 2001) indicated that the four aspects should be regarded as four separate, though interrelated, aspects, rather than as four elements of one higher order motivation variable. Therefore, in subsequent analyses, the four motivation elements were treated separately. Scale scores of both motivation instruments were transformed into new variables with a range between 0 and 1.

### *Other Variables*

It was decided to include several covariates into the analyses, based on their existence in *both* datasets. One covariate was students' average report card grade (based on a number of subject grades), which provided an indication for students' success in school and their (prior) knowledge. In the study by Brekelmans et al. (1990), a *mean report card grade* was computed by averaging report card grades of three subjects (Math, Biology, and Physics) and transforming these into a standardised variable (mean 0.0 and standard deviation 1.0). A similar variable was created in the EFL sample by averaging



(and afterwards standardising) report card grades for Dutch, Math, and English. Other variables included in the analyses were:

- student gender and the percentage of boys in the class;
- class size;
- the number of lesson minutes per week (information was gathered by asking teachers a number of questions regarding lesson length and frequency);
- school type;
- teacher gender;
- teacher experience.

Student gender and teacher gender were transformed into sets of dummy variables. For school type, two dummy variables were created, one variable contrasting intermediate general education (called HAVO) with lower general education (called MAVO or VMBO), another variable contrasting pre-university education (called VWO) with lower general education. Furthermore, it is of interest to note that in the Physics sample only half of the students of each class<sup>5</sup> were asked to complete the QTI (Questionnaire on Teacher Interaction), as this study was conducted in co-operation with a large-scale international study (SISS). In the EFL sample, all students completed the QTI.

### Respondents

For the Physics sample, 826 students (43 teachers with 43 classes from 43 schools) were included in the analyses. The EFL sample consisted of 941 students (32 teachers with 52 classes from 21 schools). Students were only included in the analyses if they had a complete set of data on all of the variables relevant in the analyses. In Table 5 both samples are described on the covariates that were included in the study.

From Table 5, it can be seen that gender distribution is roughly equal in the Physics sample. Further, it appears that most classes belong to the lower general education (MAVO) type, followed by the pre-university type (VWO). Most of the Physics teachers are males, with an average of 12 years of experience (range of experience lies between 4 and 30 years) and full

---

<sup>5</sup>From a generalisability study (Shavelson, Webb, & Burstein, 1986) it was concluded (Brekemans et al., 1990) that the QTI should be administered to at least 10 students in a class for the data to be reliable. The QTI does not need to be administered more than once per year, since interpersonal style remains relatively stable.

Table 5. Sample Characteristics for Physics and EFL.

	Covariate	Physics sample	EFL sample
Student	Student gender		
	– Male	53%	47%
	– Female	47%	53%
Class	School type <sup>a</sup>	7%	10%
	– MAVO	46%	23%
	– HAVO	18%	25%
	– VWO	29%	42%
Teacher	Teacher gender		
	– Male	97%	45%
	– Female	3%	55%
	Teaching certificate		
	– First degree	45%	43%
	– Second/third degree	54%	49%
	– No certificate	1%	8%
School	Mean experience	12 years	16 years
	Mean contract size	27 hr	24 hr
	Denomination of the school		
	– Public	21%	43%
	– Protestant	34%	19%
	– Catholic	38%	26%
	– Other	7%	12%

*Note.* <sup>a</sup>The % of combination groups of school types is displayed on the first line. In the analyses, combination groups were added to the highest school type of their combination. MAVO = lower general education; HAVO = intermediate general education; VWO = pre-university education.

appointment (the maximum number of teaching hours is 28). Teachers are equally distributed in terms of license.<sup>6</sup> Most of the classes are in Protestant or Catholic schools. The EFL sample is also equally distributed in terms of student gender and teacher license. On average, teachers have 16 years of experience and also have full appointments. However, teacher gender is more equally distributed for EFL than for Physics. Another difference between the two samples is that in the EFL sample more classes are pre-university (VWO) classes. Also, a relatively large proportion of the EFL classes go to public

<sup>6</sup>With a second degree, teachers in The Netherlands are only allowed to teach the first 3 years of secondary education; with the first degree, one may teach all education types and all years.

Table 6. Means and Standard Deviations on Outcome and Teaching Variables.

Variable	Physics sample		EFL sample	
	Mean	SD	Mean	SD
Cognitive test	7.53	1.22	7.31	1.99
Pleasure	.52	.13	.58	.09
Relevance	.57	.08	.73	.07
Confidence	.50	.10	.70	.07
Effort	.38	.09	.49	.06
Influence (DS)	.19	.36	.29	.31
Proximity (CO)	.36	.50	.53	.46

*Note.* Cognitive test scores range between 0 and 10, pleasure, relevance, confidence, and effort between 0 and 1, and influence and proximity between -3 and +3.

schools. Table 6 provides descriptive data on the outcome and interpersonal teacher behaviour variables in both samples.

Physics students on average score about 75% on the cognitive test, while they display medium subject-specific motivation in terms of pleasure, relevance, and confidence. They score below medium on effort. Furthermore, Physics teachers are perceived as somewhat dominant and co-operative, although there are large differences between teachers in terms of the latter. EFL students on average score 73% on the cognitive test and are medium motivated in terms of pleasure and effort. EFL students experience high amounts of relevance and confidence. EFL teachers are perceived as somewhat dominant and co-operative. Compared to Physics, EFL teachers are perceived as more co-operative and slightly more dominant.

### Analyses

Three multilevel models were fitted for each outcome measure in both samples, an *empty model*, a *covariate model*, and a model that included interpersonal teacher behaviour and/or interactions between interpersonal behaviour and other variables (*teaching model*).

The empty model was formulated to obtain raw estimates of variance at the different levels of the data. For the covariate model, all covariates were entered: average report card grade (as class mean and deviation of a student from the class mean), student gender, school type, teacher gender, teacher experience, percentage of boys in the class, the number of lesson minutes per week, and class size. Only variables with significant effects were retained in the analyses. Because of the absence of measures for prior achievement and

motivation, or for intelligence, the effects found in this study will probably be somewhat overestimated.

In the teaching models, teacher influence and proximity were entered, as well as interactions of these two dimensions with student or teacher characteristics (especially gender and report card grade). The interpersonal teacher behaviour variables were aggregated to the class mean and entered at the class level in the analyses. Two-level models were used in the analyses, with teacher-class combinations acting as the highest level. Analyses were conducted with MLN for Windows.

## RESULTS

In this section, we report on the outcomes of the multilevel analyses. Results are reported separately for achievement (cognitive outcomes) and pleasure, relevance, confidence, and effort (affective outcomes). Each section discusses variance distribution (research question 1), as well as (statistical) effects of interpersonal teacher behaviour (research question 2). The tables report standard regression coefficients, and in the text we report effect sizes<sup>7</sup> of variables, in order to compare them with each other (Appendix A contains all the effect sizes of the teaching models).

### Achievement

Outcomes of the empty model show that for EFL a large amount of variance in achievement is located at the teacher-class level, and that this amount is moderate for Physics (45.3%<sup>8</sup> vs. 20.0%). These differences in variance distribution between the two samples remain after correction of achievement for student and class characteristics, leaving 4.9% of the unexplained variance at the teacher-class level for the Physics sample and 18.4% for the EFL sample (see Table 7). Thus, after correction for covariates, differences between teacher-class combinations are more than twice as large in the EFL sample as in the Physics sample.

---

<sup>7</sup>The effect size is reported in terms of the standardised regression coefficient, which can be obtained with the formula (Cohen, 1969): standardised coefficient = coefficient \* (standard deviation independent variable/standard deviation dependent variable).

<sup>8</sup>This percentage of class-level variance is very high, compared to other studies (e.g., Reynolds, 1995).

Table 7. Achievement, Significant (at  $\alpha = .05$ ) Regression Coefficients and Variance Components for the Physics and EFL Sample.

		Empty model		Covariate model		Teaching model	
		Physics	EFL	Physics	EFL	Physics	EFL
Student	Constant	7.560	7.324	6.669	4.933	6.593	4.958
	Gender			-.305	-	-.304	-
	Rep crd-dev			.352	.453	.352	.346
Class	HAVO			.537	.994	.609	1.004
	VWO			1.081	1.366	1.110	1.417
	Rep crd-av			.279	.891	.345	.868
	Percentage boys			-	.029	-	.029
	DS (influence)					.344	-
	CO (proximity)					-	-
Variance	Rep crd * CO					-	.346
	Student	80.0	54.7	68.8	51.3	68.6	50.7
	Class	20.0	45.3	3.4	11.5	2.9	11.1
	Explained	-	-	27.8	37.2	28.5	38.2
-2 * log(like)		2598.1	3579.5	2422.2	3463.6	2415.4	3451.4
Difference		-	-	175.66 (5)	115.90 (5)	7.07 (1)	12.17 (2)
log (df)							

Note. Rep crd = mean report card grade; gender = student gender (male = baseline); HAVO = school type HAVO (MAVO = baseline); VWO = school type VWO (MAVO = baseline); dev = individual difference from class mean; av = class mean.

Interpersonal teacher behaviour only is significantly related with performance in the Physics sample, but has no significant association with performance in the EFL sample. Higher dominance relates to higher cognitive outcomes for Physics. The effect size for Influence can be regarded as moderate (see Appendix A). A difference of one standard deviation in Influence corresponds to half of the differences between school types, and to half of the difference in achievement between boys and girls. No association was found between Physics teachers' Proximity and their students' cognitive outcomes. For the EFL sample, no significant relations were found. However, an interaction was found between Proximity and individual mean report card grade. Thus, EFL students with co-operative teachers according to mean student perceptions and high mean report card grades outperformed other students. As can be expected, percentages of explained variance at the teacher-class level by interpersonal teacher behaviour variables or their interactions are higher in the Physics sample (14.7%) than in the EFL sample (3.5%).

### **Pleasure**

As can be seen in Table 8, many differences exist between the Physics sample and EFL sample with respect to pleasure. Differences between classes in pleasure are much more important in the Physics sample than in the EFL sample (31.6% of the variance is located at the teacher-class level in the Physics sample vs. 13.5% in the EFL sample), even after correction for covariates (25.0% of the unexplained variance vs. 11.1%) and entering of teacher behaviour variables (8.3% of the unexplained variance vs. 5.9%).

When looking at teacher behaviour, teacher Proximity is important for pleasure in both the Physics and EFL samples, although its association is slightly stronger in the Physics sample (see Appendix A). A gain of one standard deviation in Proximity is three times stronger than the difference in pleasure between boys and girls in both samples. Apart from that, Influence is also important for pleasure in the Physics sample, but not for pleasure in the EFL sample. In terms of standardised effects (Appendix A), the effect of Influence is only less than a quarter of that of Proximity. Interpersonal behaviour explains two thirds of the class-level variance for Physics and 40% for EFL.

### **Relevance**

For relevance, a reversed picture in terms of variance distribution is found as for pleasure (see Table 9). Differences between teacher-class combinations in

Table 8. Pleasure, Significant Regression Coefficients and Variance Components for the Physics and EFL Sample.

	Empty model		Covariate model		Teaching model	
	Physics	EFL	Physics	EFL	Physics	EFL
Student level						
Constant	.517	.575	.467	.698	.457	.629
Gender			-.069	.019	-.068	.021
Rep crd-dev			.047	.020	.047	.019
Class level						
Percentage boys			.002	-	-	-
Lesson minutes			-	-.001	-	-.001
DS (influence)					.047	-
CO (proximity)					.173	.112
Variance						
Student	68.4	86.5	66.0	86.5	66.0	86.5
Class	31.6	13.5	22.0	10.8	6.0	5.4
Explained	-	-	12.0	2.7	28.0	8.1
-2 * log(like)	-267.1	-496.4	-377.6	-511.0	-424.2	-537.3
Difference log (df)	-	-	110.46 (3)	14.54 (3)	46.64 (2)	26.3 (1)

Note. Rep crd = mean report card grade; gender = student gender (male = baseline); lesson minutes = number of lesson minutes per week; dev = individual difference from class mean; av = class mean.

Table 9. Relevance, Significant Regression Coefficients and Variance Components for the Physics and EFL Sample.

	Empty model		Covariate model		Teaching model	
	Physics	EFL	Physics	EFL	Physics	EFL
Student level						
Constant	.575	.736	.531	.625	.526	.616
Gender			-.047	-	-.048	-
Rep crd-dev			.046	.012	.046	.012
Class level						
Rep crd grade-av			.043	.030	.038	.031
Percentage boys			.001	.002	.001	.002
DS (influence)					.073	-
CO (proximity)					.045	-
Variance						
Student	90.9	85.7	80.6	85.7	80.6	85.7
Class	9.1	14.3	5.6	9.5	4.8	9.5
Explained	-	-	13.8	4.8	14.6	4.8
-2 * log(like)	-439.8	-1044.0	-536.5	-1067.6	-557.1	-1067.6
Difference log (df)	-	-	96.67 (4)	23.58 (3)	20.61 (2)	0 (0)

Note. Rep crd = mean report card grade; gender = student gender (male = baseline); dev = individual difference from class mean; av = class mean.



perceived relevance are somewhat larger in the EFL sample than in the Physics sample (14.3% vs. 9.1%) and these differences remain in the covariate and teaching models.

As can be seen from the empty model, students on average rate relevance higher in the EFL sample than in the Physics sample. Interpersonal teacher behaviour only has an association with perceived relevance in the Physics sample, but not in the EFL sample. For Physics teachers, Influence and Proximity both have a positive association with the level of student relevance. The effect of Influence is slightly larger than that of Proximity in terms of its standardised effect size for relevance (Appendix A). A gain of one standard deviation in Proximity results in a difference that is almost equal to the difference in relevance between boys and girls. Its effect is similar to that of average report card grade or gender distribution within a class. Interpersonal teacher behaviour explains a fair amount of variance at the teacher-class level (33%) in Physics.

### Confidence

Similar distributions of variance can be found for the two samples with respect to confidence. Although amounts of variance in confidence at the teacher-class level are twice as large in the Physics sample as in the EFL sample (11.5 vs. 5.4%) – and remain so after correction for covariates and interpersonal teacher behaviour – similar variables account for variance (see Table 10).

On average, students in the EFL sample have more confidence than those in the Physics sample (see empty model of Table 10). In both samples, Influence and Proximity are significantly associated with confidence. More teacher dominance is associated with *less* student confidence, while more teacher Proximity is associated with *more* student confidence. A gain of one standard deviation in Influence results in a difference in confidence that is twice as large in the Physics sample as in the EFL sample (Appendix A). It should be noted that the effect size for Proximity is nearly equal in both samples. Also, the effect for Proximity is similar to that of student gender, but much smaller than the effect of report card grade. Thus, the teacher's behaviour is stronger related with the confidence of students in the Physics sample than in the EFL sample. An additional interaction effect between average (class) mean report card grade and Influence was found in the EFL sample. Students in EFL classes with higher average mean report card grades and more dominant teachers have less confidence than students in other classes. Another interaction effect was found between student gender and Proximity in the Physics

Table 10. Confidence, Significant Regression Coefficients and Variance Components for the Physics and EFL Sample.

	Empty model			Covariate model			Teaching model		
	Physics	EFL		Physics	EFL		Physics	EFL	
Student level									
Constant	.504	.703		.542	.722		.558	.722	
Gender				-.076	-.037		-.099	-.035	
Rep crd-dev				.093	.040		.094	.060	
Class level									
Rep crd-av				-	.029		-	.034	
DS (influence)									
CO (proximity)							-.148	-.059	
Rep crd * DS							.051	.030	
Gender * CO							-	-.068	
Variance							.077	-	
Student	88.5	94.6		65.4	89.2		63.5	89.2	
Class	11.5	5.4		11.5	5.4		9.6	2.7	
Explained	-	-		23.1	5.4		26.9	8.1	
-2 * log(like)	-147.8	-451.5		-385.8	-489.9		-403.4	-508.9	
Difference log (df)	-	-		238.02 (2)	38.43 (3)		17.55 (3)	19.02 (3)	

Note. Rep crd = mean report card grade; gender = student gender (male = baseline); dev = individual difference from class mean; av = class mean.

Table 11. Effort, Significant Regression Coefficients and Variance Components for the Physics and EFL Sample.

	Empty model		Covariate model		Teaching model	
	Physics	EFL	Physics	EFL	Physics	EFL
Student level						
Constant	.376	.488	.342	.402	.402	.532
Gender			-.113	-	-.116	-
Rep crd-dev			.078	-	.068	-
Class level						
Lesson minutes			-	-.001	-	-.001
DS (influence)					.068	.042
CO (proximity)					.050	.066
Variance						
Student	92.6	92.0	68.5	92.0	68.5	92.0
Class	7.4	8.0	3.7	8.0	3.7	4.0
Explained	-	-	27.8	-	27.8	4.0
-2 * log(like)	-88.7	-843.3	-341.7	-847.5	-349.1	-862.8
Difference	-	-	253.05 (2)	4.26 (1)	7.31 (2)	15.31 (2)
log (df)						

Note. Gender = student gender (male = baseline); rep crd = mean report card grade; lesson minutes = number of lesson minutes per week; dev = individual difference from class mean; av = class mean.

sample. This interaction indicates that girls in Physics classes with more co-operative teachers have more confidence than boys. On the whole, more teacher-class level variance is explained by interpersonal teacher behaviour in the EFL sample (50%) than in the Physics sample (about 15%).

### **Effort**

Similar amounts of teacher-class level variance in effort are found in the Physics and EFL samples (see Table 11). In both samples, these amounts of unexplained variance are about 7%, and about 4.5% after correction for covariates and teacher behaviour. As can be seen in Table 11, ratings for effort are on average higher in the EFL sample than in the Physics sample.

Teacher Proximity and Influence are both positively associated with effort in the Physics sample as well as in the EFL sample. The effect of Influence is similar in the Physics and EFL samples (Appendix A), while the effect of Proximity is almost twice as large in the EFL sample as in the Physics sample. Although the associations with interpersonal behaviour are similar in both samples, more variance is explained at the teacher-class level in the EFL sample than in the Physics sample. In the EFL sample, 50% of the teacher-class variance is explained by introducing Influence and Proximity in the models, whereas hardly any variance is explained in the Physics sample.

## **DISCUSSION AND CONCLUSION**

The results of this study show that variance in outcomes is distributed equally over the student and teacher-class levels for both samples with respect to effort, while the other affective outcomes and achievement show differences in variance decomposition. More variance is located at the teacher-class level in the Physics sample than in the EFL sample with respect to pleasure and confidence, and smaller amounts at the teacher-class level are found for achievement and relevance. This means that teachers in the Physics sample seem to make a bigger difference than teachers in the EFL sample for their students' pleasure and confidence, whereas they seem to have less impact on achievement and relevance.

Relations of interpersonal teacher behaviour are different with cognitive and affective outcomes: Proximity has a strong positive relation with students' subject-specific motivation, but it has no association with students' cognitive test scores. Moreover, the associations between interpersonal behaviour and

cognitive outcomes are much smaller – both in terms of effect sizes and percentages of explained variance – than they are with students' subject-specific motivation. This finding corresponds with earlier Dutch research (Brekelmans et al., 1990, 2002; Den Brok, 2001). The associations show that students' perceptions of the cooperativeness of their teachers are very important for their own motivation, though less important for achievement.

Effects of interpersonal teacher behaviour are also different in the Physics and EFL samples: whereas teacher Influence has no association with cognitive test scores in the EFL sample, it is (slightly) positively related to test scores in the Physics sample. Also, Influence is related positively to pleasure, relevance, and effort in the Physics sample, whereas it is only related to effort in the EFL sample. Influence is negatively related to confidence in both the Physics and EFL samples, though stronger in the latter sample. Finally, interpersonal behaviour explains more variance in effort and confidence in the EFL sample, whereas it explains more variance in pleasure, relevance, and test scores in the Physics sample. It is possible that these differences to some degree reflect differences in nature between subjects. English teachers may have less effect on their students' achievement, pleasure, or relevance compared to Physics teachers, because students often have much opportunity to practice the language outside the classroom, for example by watching television at home, surfing the internet, or listening to the radio. Students often have less opportunity to occupy themselves with the content of Physics after school. Moreover, in English lessons, teachers often use tasks consisting of many small subtasks requesting direct correction (for example, speaking or writing sentences or words), which may have led to higher dominance scores and less variance in dominance for the EFL sample as compared to the Physics sample. Of course, the above explanations remain speculative, as they have not been investigated explicitly.

The finding that certain teacher behaviour may have different effects in one subject sample as compared to another, or for one outcome measure as compared to another, is also in line with earlier findings by effectiveness researchers (Luyten, 1994, 1998; Reynolds, 1995). It may provide some support for the assertion that instability of school effects in terms of subjects and/or outcome types probably finds its origin in mechanisms that are located at the teacher or teacher-class level.

Unfortunately, the study suffered from a number of limitations. First, the two samples studied were completely independent: for example, they consisted of different schools, teachers, and students. This prevented us from *directly*

investigating stability of interpersonal effectiveness *across* subjects. Future research, investigating the relationship between interpersonal behaviour and student outcomes for different subjects within the same school or with teachers teaching multiple subjects, could shed light on this issue. Another limitation was the fact that models mainly focused on associations between teaching and student outcomes. The models lacked prior achievement or motivation and therefore the "added value" of the teacher (interpersonal behaviour) was not investigated. The number of covariates included in this study was limited, because the two samples shared only a small part of their (original) variables. This was caused by the fact that the samples were gathered with an interval of more than 10 years between them and were both part of larger studies, initiated for different reasons. The lack of certain covariates, such as prior outcomes, may have led to overestimation of the strength of relationships between interpersonal teacher behaviour and outcomes. Preliminary analyses on the EFL data including such covariates seem to confirm this: The effect of interpersonal behaviour reduces and becomes nonsignificant for achievement, whereas it slightly reduces for affective outcomes (though remaining significant) (Den Brok, 2001). A third limitation was the fact that interpersonal behaviour was the only teaching variable included in the models. Again, this may have led to overestimation of the effects of interpersonal teacher behaviour. However, preliminary analyses again indicate that this behaviour remains important after other behaviours have been added to the models (Den Brok, 2001), a finding that is in line with prior research (Goh, 1994; Henderson, 1995; Rawnsley, 1997). Fourth, whereas the study investigated associations between (students' perceptions of) teacher behaviour and student outcomes, these investigations did not include structural models. Future research could employ such structural analyses, thereby establishing the strength and directions of relationships between variables. Fifth, the number of teachers included in both samples is relatively small, and prevents us from distinguishing more than two levels in the analyses. Research with larger samples and more levels of analysis is needed in the future. Such research could also shed light on the stability of teaching and teaching effects across classes of the same teacher, as well as differences in teaching effects across schools. Moreover, other subjects could be included in such samples.

This study measured teacher behaviour in terms of *students' perceptions* of their teachers interpersonal behaviour. Brekelmans and colleagues (1990) investigated links between *teachers' (self-)perceptions* of interpersonal behaviour and student outcomes for Physics, but hardly found any significant

associations. Their study also showed considerable differences between students' and teachers' perceptions of interpersonal behaviour, a finding that has been replicated in many other studies using the QTI, as well as with other teacher behaviour questionnaires (Den Brok, Bergen, & Brekelmans, 2003). Future research could investigate if a link exists between teachers' (self-) perceptions of their interpersonal behaviour and student outcomes and verify the findings by Brekelmans et al. (1990) for other subjects and samples.

Of course, students' perceptions could be linked to observational data typically gathered in effectiveness research, such as learning time, opportunity to learn, direct instruction, et cetera. Such research may help to uncover interesting associations and determine the nature and importance of the teacher-student interpersonal relationship for other effectiveness factors.

The outcomes of our study show important associations between interpersonal teacher behaviour and student outcomes: it explains between 14.7 and 67% of the class-level variance in a Physics sample and between 3.5 and 50% of the class-level variance in an EFL sample. This suggests that it may be a very relevant variable for effectiveness researchers. There may be an additional reason for effectiveness researchers to include interpersonal teacher behaviour in their models. Research from cross-national studies (Den Brok, Fisher, et al., 2003; Wubbels & Levy, 1991) and cross-cultural studies (Levy et al., 2003) using the Questionnaire of Teacher Interaction (see also Den Brok, Levy, Rodriguez, & Wubbels, 2002) indicates that the instrument and model are cross-culturally valid. This opens up opportunities for researchers to use it in large-scale international effectiveness studies, like TIMMS.

### ACKNOWLEDGEMENT

The authors would like to thank the editors of *School Effectiveness and School Improvement* and three anonymous reviewers for their comments on an earlier version of this manuscript. This study was supported with a grant of the Dutch Organisation for Scientific Research (NWO, project 575-90-804).

### REFERENCES

- Blackburn, R., & Renwick, S.J. (1996). Rating scales for measuring the interpersonal circle in forensic psychiatric patients. *Psychological Assessment*, 8(1), 76-84.
- Blumenfeld, P.C., & Meece, J.L. (1985). Life in classrooms revisited. *Theory Into Practice*, 24, 50-56.

- Brekelmans, M., Sleegers, P., & Fraser, B.J. (2000). Teaching for active learning. In P.R.J. Simons, J.L. van der Linden, & T. Duffy (Eds.), *New learning* (pp. 227–242). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Brekelmans, M., Wubbels, Th., & Créton, H.A. (1990). A study of student perceptions of physics teacher behavior. *Journal of Research in Science Teaching*, 27, 335–350.
- Brekelmans, M., Wubbels, Th., & Den Brok, P. (2002). Teacher experience and the teacher-student relationship in the classroom environment. In S.C. Goh & M.S. Khine (Eds.), *Studies in educational learning environments: An international perspective* (pp. 73–100). Singapore: New World Scientific.
- Brophy, J.E., & Good, T.L. (1986). Teacher behaviour and student achievement. In M.C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 328–375). New York: Macmillan.
- Brown, R. (1965). *Social psychology*. London: Collier-McMillan.
- Clément, R., Dörnyei, Z., & Noels, K.A. (1994). Motivation, self-confidence, and group cohesion in the foreign language classroom. *Language Learning*, 44, 417–448.
- Cohen, J. (1969). *Statistical power analysis for the behavioural sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Creemers, B.P.M. (1994). *The effective classroom*. London: Cassell.
- Creemers, B.P.M., & Reezigt, G.J. (1996). *Common elements in theories about effective instruction*. Groningen, The Netherlands: ICO.
- Creemers, B.P.M., & Scheerens, J. (1994). Developments in the educational effectiveness research programme. *International Journal of Educational Research*, 21(2), 125–140.
- d'Agostino, J.V. (2000). Instructional and school effects on students' longitudinal reading and mathematics achievements. *School Effectiveness and School Improvement*, 11, 197–235.
- Den Brok, P. (2001). *Teaching and student outcomes*. Utrecht, The Netherlands: WCC.
- Den Brok, P., Bergen, T., & Brekelmans, M. (2003). *Comparing teachers' and students' perceptions of instructional behaviour in Dutch secondary education*. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago.
- Den Brok, P., Fisher, D.L., Brekelmans, M., Rickards, T., Wubbels, Th., Levy, J., & Waldrup, B. (2003, April). *Students' perceptions of secondary science teachers' interpersonal style in six countries: A study on the validity of the Questionnaire on Teacher Interaction (QTI)*. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago.
- Den Brok, P., Levy, J., Rodriguez, R., & Wubbels, T. (2002). Perceptions of Asian-American and Hispanic-American teachers and their students on interpersonal communication style. *Teaching and Teacher Education*, 18, 447–467.
- Den Brok, P., Levy, J., Wubbels, T., & Rodriguez, M. (2003). Cultural influences on students' perceptions of videotaped lessons. *International Journal of Intercultural Relations*, 27, 268–288.
- Doyle, W. (1979). Classroom effects. *Theory Into Practice*, 18, 138–144.
- Doyle, W. (1986). Classroom organization and management. In M.C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 392–431). New York: Macmillan.
- Driscoll, A., Peterson, K., Crow, H., & Larson, B. (1985). Student reports for primary teacher evaluation. *Educational Research Quarterly*, 9(3), 43–50.
- Dunkin, M., & Biddle, B. (1974). *The study of teaching*. New York: Holt, Rinehart & Winston.
- Evans, H.M. (1998). *A study of students' cultural background and teacher-student interpersonal behavior in secondary science classrooms in Australia*. Unpublished



- doctoral dissertation, Science and Mathematics Education Centre, Curtin University, Perth, Australia.
- Fabrigar, L.R., Visser, P.S., & Browne, M.W. (1997). Conceptual and methodological issues in testing the circumplex structure of data in personality and social psychology. *Personality and Social Psychology Review*, 1, 184–203.
- Fisher, D.L., Fraser, B.J., & Wubbels, Th. (1992). *Teacher communication style and school environment*. Paper presented at the 1992 ECER conference, Enschede, The Netherlands.
- Foa, U.G. (1961). Convergence in the analysis of the structure of interpersonal behavior. *Psychological Review*, 68, 341–353.
- Fraser, B.J. (1994). Research on classroom and school climate. In D. Gabel (Ed.), *Handbook of research in science teaching and learning* (pp. 483–541). New York: Macmillan.
- Fraser, B.J. (1998). Science learning environments: Assessment, effects and determinants. In B.J. Fraser & K.G. Tobin (Eds.), *International handbook of science education* (Part 1, pp. 527–564). London: Kluwer Academic Publishers.
- Gaines, S.O., Panter, A.T., Lyde, M.D., Steers, W.N., Rusbult, C.E., Cox, C.L., & Wexler, M.O. (1997). Evaluating the circumplexity of interpersonal traits and the manifestation of interpersonal traits in interpersonal trust. *Journal of Personality and Social Psychology*, 73, 610–623.
- Gardner, R.C., & Lambert, W.E. (1972). *Attitudes and motivation in second language learning*. Rowley, MA: Newbury House.
- Gardner, R.C., & Macintyre, P.D. (1993). On the measurement of affective variables in second language learning. *Language Learning*, 43, 157–194.
- Goh, S.C. (1994). *Interpersonal teacher behavior, classroom climate and student outcomes in primary mathematics classes in Singapore*. Unpublished doctoral dissertation, Science and Mathematics Education Centre, Curtin University of Technology, Perth, Australia.
- Grisay, A. (1996). *Evolution des acquis cognitifs et socio-affectives des élèves au cours des années de collège* [Development of cognitive and socio-affective learning of pupils in secondary education]. Liège, Belgium: Université de Liège.
- Gurtman, M.B., & Pincus, A.L. (2000). Interpersonal adjective scales: Confirmation of circumplex structure from multiple perspectives. *Personality and Social Psychology Bulletin*, 26, 374–384.
- Henderson, D.G. (1995). *A study of the classroom and laboratory environments and student attitude and achievement in senior secondary biology classes*. Unpublished doctoral dissertation, Science and Mathematics Education Centre, Curtin University, Perth, Australia.
- Hill, P.W., & Rowe, K.J. (1996). Multilevel modelling in school effectiveness research. *School Effectiveness and School Improvement*, 7, 1–34.
- Hox, J.J. (1995). *Applied multilevel analysis*. Amsterdam: TT Publications.
- Kyriakides, L., Gagatsis, A., & Campbell, R.J. (2000). The significance of the classroom effect in primary schools: An application of Creemers comprehensive model of educational effectiveness. *School Effectiveness and School Improvement*, 11, 501–529.
- Leary, T. (1957). *An interpersonal diagnosis of personality*. New York: Ronald Press Company.
- Lee, O. (1995). Subject matter knowledge, classroom management, and instructional practices in middle school science classrooms. *Research in Science Teaching*, 32(4), 423–440.
- Levy, J., Den Brok, P., Wubbels, Th., & Brekelmans, M. (2003). Significant variables in students' perceptions of teacher interpersonal communication styles. *Learning Environments Research*, 6, 5–36.

- Levy, J., Wubbels, Th., & Brekelmans, M. (1992). Student and teacher characteristics and perceptions of teacher communication style. *Journal of Classroom Interaction*, 27, 23–29.
- Lonner, W.J. (1980). The search for psychological universals. In H.C. Triandis & W.W. Lambert (Eds.), *Handbook of cross cultural psychology* (Vol. 1, pp. 143–204). Boston: Allyn and Bacon.
- Lowyck, J. (1994). Teaching effectiveness: An overview of studies. *Tijdschrift voor Onderwijsresearch*, 19, 17–25.
- Luyten, H. (1994). Stability of school effects in Dutch secondary education: The impact of variance across subjects and years. *International Journal of Educational Research*, 21, 197–216.
- Luyten, H. (1998). School effectiveness and student achievement, consistent across subjects? Evidence from Dutch elementary and secondary education. *Educational Research and Evaluation*, 4, 281–306.
- Luyten, H., & De Jong, R. (1998). Parallel classes: Differences and similarities. Teacher effects and school effects in secondary schools. *School Effectiveness and School Improvement*, 9, 437–473.
- Mak, A.M. (2001). *Rating of teacher behaviour by primary school students: Are they up to it?* Unpublished Masters Thesis, Twente University, Enschede, The Netherlands.
- Mandeville, G.K., & Anderson, L.W. (1987). The stability of school effectiveness indices across grade levels and subject area. *Journal of Educational Measurement*, 24, 203–216.
- Moos, R.H. (1979). *Evaluating educational environments: Procedures, measures, findings and policy implications*. San Francisco: Jossey-Bass.
- Murray, H.A. (1938). *Explorations in personality*. New York: Oxford University Press.
- Muthén, B. (1994). Multilevel covariance structure analysis. *Sociological Methods and Research*, 22, 338–354.
- Muthén, L.K., & Muthén, B.O. (1999). *Mplus user's guide*. Los Angeles: Authors.
- Peterson, K., & Stevens, D. (1988). Student reports for school teacher evaluation. *Journal of Personnel Evaluation in Education*, 2, 19–31.
- Rawnsley, D.G. (1997). *Associations between classroom learning environments, teacher interpersonal behavior and student outcomes in secondary Mathematics classrooms*. Unpublished doctoral dissertation, Science and Mathematics Education Centre, Curtin University, Perth, Australia.
- Reynolds, D. (1995). The effective school: An inaugural lecture. *Evaluation and Research in Education*, 9, 57–73.
- Sandal, O., Wold, B., & Bronis, M. (1999). Relationship between students' perceptions of school environment, their satisfaction with school and perceived academic achievement: An international study. *School Effectiveness and School Improvement*, 10, 296–320.
- Scheerens, J. (1993). Basic school effectiveness research: Items for a research agenda. *School Effectiveness and School Improvement*, 4, 17–36.
- Scheerens, J., & Bosker, R. (1997). *The foundations of educational effectiveness*. Oxford: Pergamon.
- Scheerens, J., & Creemers, B.P.M. (1996). School effectiveness in the Netherlands: The modest influence of a research programme. *School Effectiveness and School Improvement*, 7, 181–195.
- Scriven, M. (1994). Using student ratings in teacher evaluation. *Evaluation Perspectives*, 4(1), 3–6.

- Shavelson, R.J., Webb, N.W., & Burstein, L. (1986). Measurement of teaching. In M.C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 50–91). New York: Macmillan.
- Shuell, T.J. (1996). Teaching and learning in a classroom context. In D.C. Berliner & R.C. Calfee (Eds.), *Handbook of educational psychology* (pp. 726–764). New York: Macmillan.
- Shulman, L. (1986). Paradigms and research programs in the study of teaching: A contemporary perspective. In M.C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 3–36). New York: Macmillan.
- Slater, P.E. (1962). Parental behavior and the personality of the child. *Journal of Genetical Psychology*, 101, 53–68.
- Snijders, T.A.B., & Bosker, R.J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modelling*. London: Sage.
- Strack, S. (1996). Special series: Interpersonal theory and the interpersonal circumplex: Timothy Leary's Legacy. *Journal of Personality Assessment*, 66, 211–307.
- Taba, H., Tylor, R., & Smith, B.O. (1998). *The predictive validity of student evaluations in the identification of meritorious teachers*. Research Report. Washington, DC: John Jones School District.
- Van Amelsvoort, J. (1999). *Perspective on instruction, motivation and selfregulation*. Dissertation. Nijmegen, The Netherlands: University Press.
- Van Damme, J., De Fraine, B., Van Landeghem, G., Opdenakker, M.-C., & Onghena, P. (2002). A new study on educational effectiveness in Secondary schools in Flanders: An introduction. *School Effectiveness and School Improvement*, 13, 383–398.
- Van de Grift, W., Houtveen, T., & Vermeulen, C. (1997). Instructional climate in Dutch Secondary education. *School Effectiveness and School Improvement*, 8, 449–462.
- Walberg, H.J. (1979). *Educational environments and effects: Evaluation, policy, and productivity*. Berkeley, CA: McCutchan.
- Watzlawick, P., Beavin, J.H., & Jackson, D. (1967). *The pragmatics of human communication*. New York: Norton.
- Willms, J.D., & Raudenbush, S.W. (1989). A longitudinal hierarchical linear model for estimating school effects and their stability. *Journal of Educational Measurement*, 26, 209–231.
- Wubbels, T., & Levy, J. (1991). A comparison of interpersonal behavior of Dutch and American teachers. *International Journal of Intercultural Relations*, 15, 1–18.
- Wubbels, T., & Levy, J. (1993). *Do you know what you look like?* London: Falmer Press.
- Wubbels, Th., & Brekelmans, M. (1997). A comparison of student perceptions of Dutch Physics teachers' interpersonal behaviour and their educational opinions in 1984 and 1993. *Journal of Research in Science Teaching*, 5, 447–466.
- Wubbels, Th., & Brekelmans, M. (1998). The teacher factor in the social climate of the classroom. In B.J. Fraser & K.G. Tobin (Eds.), *International handbook of Science education, part one* (pp. 565–580). London: Kluwer Academic Publishers.
- Wubbels, Th., Créton, H.A., & Hooymayers, H.P. (1985). *Discipline problems of beginning teachers, interactional teacher behavior mapped out*. Abstracted in Resources in Education, 20, 12, p. 153, ERIC document 260040.

## APPENDIX A

Effect sizes (Standardized Effects) of Variables in the Teaching Models.

Variable	Cognitive test		Pleasure		Relevance		Confidence		Effort	
	Phys	EFL	Phys	EFL	Phys	EFL	Phys	EFL	Phys	EFL
Gender	-.125		-.262	.117	-.300		-.495	-.250	-.644	
Rep crd-dev	.262	.143	.328	.173	.521	.141	.852	.703	.685	
HAVO	.215	.227								
VWO	.446	.356								
Rep crd-av	.122	.244			.204	.248		.272		
Percentage boys		.186			.228	.364				
Lesson minutes				-.277						-.415
DS (influence)	.103		.132		.332		-.539	-.261	.275	.217
CO (proximity)			.671	.577	.284		.257	.199	.280	.510
Rep crd * DS								-.169		
Rep crd * CO		.045								
Gender * CO							.194			

*Note.* Rep crd = mean report card grade; gender = student gender (male = baseline); HAVO = school type HAVO (MAVO = baseline); VWO = school type VWO (baseline = MAVO); lesson minutes = number of lesson minutes per week; dev = individual difference from class mean; av = class mean.

Copyright of School Effectiveness & School Improvement is the property of Taylor & Francis Ltd and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.