British Journal of Educational Psychology (2002), 72, 467–482

www.bps.org.uk

The dimensionality of speaking and writing: A multilevel factor analysis of situational, task and school effects

Hans Kuhlemeier[1*], Huub van den Bergh[2] and Gert Rijlaarsdam[3]

[1]National Institute for Educational Measurement (Cito), Arnhem, The Netherlands
[2]Centre for Language and Communication, University of Utrecht, The Netherlands
[3]Graduate School of Teaching and Learning, University of Amsterdam, The Netherlands

Background. This article addresses the question whether communication apprehension (CA) should be regarded as a unidimensional or alternatively as a multidimensional construct. The answer is not only interesting from a theoretical point of view. There might also be practical implications for the treatment of CA. If CA were to appear to be unidimensional and a student's level of CA were to be the same across situations and tasks, there would be no need to tailor the treatment to particular speaking situations or specific writing problems. If, however, CA appeared to be multidimensional and a student might have a variety of different levels and types of CA, one type of treatment might be more effective for one student than for another one.

Aim. To examine the effects of situational, task and school effects on speaking and writing apprehension.

Sample. Use was made of the dataset of the 1987–1988 National Assessment of Language Performance in the Netherlands. The nationally representative sample consisted of 1448 students from 184 secondary schools; 52% of the students were boys and 48% were girls; the mean age of the students was 15 years 6 months.

Method. Speaking and writing apprehension were measured by means of self-report measures in grade 9. Multilevel factor analysis (MLFA) was used to determine the dimensionality of the measurement of speaking and writing apprehension.

Results and conclusions. First, all seven speaking situations and three out of four writing problems could be distinguished empirically. Speaking and writing apprehension are clearly multidimensional constructs that depend on the speaking situation and the writing task. Second, correlations between speaking and writing apprehension were

* *Requests for reprints should be addressed to Dr. H. Kuhlemeier, National Institute for Educational Measurement (Cito), PO Box 1034, 6801 MG Arnhem, The Netherlands (e-mail: hans.kuhlemeier@citogroep.nl).*

rather low. Speaking and writing apprehension seem to represent skill-specific constructs, which cannot be considered as equivalent forms of communication apprehension. Third, differences between schools in the level of speaking and writing apprehension were very small compared to measurements of speaking and writing performance.

Most people experience some form of communication apprehension (CA). High CA is not only a major source of emotional and psychological distress but it may also have pervasive educational and social consequences (McCroskey, 1977; McCroskey, Daly, & Sorensen, 1976). Compared to low CA students, high CA students have lower self-esteem, are perceived as less attractive and credible by peers, have fewer dates with members of the opposite sex, are four times as likely to sit outside the high interaction area in the classroom, and participate less in small group interaction. Furthermore, high CA students, while not being less intelligent, have lower standardised achievement test scores throughout elementary and secondary education and obtain lower grades. Moreover, high CA students tend to avoid high status occupations that have high communication requirements. There is hardly a language teacher who has not recognised the problem and would not be willing to help students to overcome high CA. Currently a number of effective treatments are available, including skills training, systematic desensitisation, and cognitive modelling (Allen, Hunter, & Donahue, 1989; Richmond & McCroskey, 1998; Witworth & Cochran, 1996).

CA is conceptualised as a multidimensional construct. Originally CA was narrowly defined as a broadly based anxiety related to oral communication (McCroskey, 1970). CA has later been broadened explicitly to encompass other constructs such as writing apprehension (Daly & Miller, 1975) and singing apprehension (Andersen, Andersen, & Garrison, 1978). Only recently CA has been expanded even further to include two new constructs: intercultural CA and interethnic CA. These new constructs are conceptualised as 'the fear or anxiety associated with either real or anticipated interaction with people of different groups, especially cultural and ethnic and/or racial groups' (Neuliep & McCroskey, 1997, p. 385).

Research into the relationships between the constructs of CA has yielded contradictory evidence. On the one hand, speaking apprehension is found to be highly correlated with intercultural CA (Neuliep & McCroskey, 1997). This finding is consistent with the view of CA encompassing several highly related but discrete constructs. Contrary to this view there is some research evidence that speaking apprehension (SA) and writing apprehension (WA) are not substantially related. Klopf & Cambra (1979) found SA and WA to be totally unrelated, while McCroskey (1977) reported rather low correlations ranging between .30 and .40. This would suggest that SA and WA represent separate, skill-specific constructs, which can hardly be considered as equivalent forms of communication apprehension. In this article the relationship between oral and written CA is examined in detail.

CA and its subconstructs may be conceptualised at different levels of generalisation across contexts, situations, tasks and time (cf. Spielberger, 1966). While early work conceptualised CA primarily as a highly generalisable and stable personality trait (McCroskey, 1970), more recent conceptualisations posit CA on a continuum ranging from an extreme trait pole to an extreme state pole (McCroskey, 1984). On this continuum four types of CA are distinguished: trait-like CA, context-based CA, audience-based CA, and situational CA (McCroskey, Richmond, & Davis, 1986). Only trait-like CA

and context-based CA will be considered in this article. The trait-like view defines CA as 'a relatively enduring, personality-based orientation toward a given mode of communication across a wide variety of contexts' (McCroskey, 1984, p. 16). It predicts a person having fear of communication in one context having the same level of fear in any other context. The context-based view of CA states that students may encounter high levels of CA in one type of context, while having little or no apprehension in another context. For instance, speaking apprehension (SA) may vary as a function of the size, composition and status of the audience (Daly & Buss, 1984; Geen, 1991). A student may feel very anxious about giving a talk in front of the whole class, while experiencing participating in a small-group discussion as rather pleasant. Analogously, writing apprehension (WA) may not always be consistent across different elements of the writing task. When writing an essay, a student may not have any problem with the mechanics of writing, while problems in generating relevant content may cause long-lasting blocks.

The question of whether SA and WA should be regarded as unidimensional or alternatively as multidimensional subconstructs of CA has been addressed in several studies. Research into the influence of situational and task factors on the level of CA, SA and WA has yielded contradictory results. The most frequently used and studied instrument for measuring trait CA is probably the Personal Report of Communication Apprehension (PRCA) (McCroskey, 1970). Originally the PRCA was developed from a unidimensional trait perspective (McCroskey, 1970). A later version of the instrument – the PRCA24 – claimed to measure unidimensional trait CA as well as multidimensional context-based CA. The newest version consists of four subtests measuring CA in public speaking situations, group meetings, small groups, and two-person conversation. McCroskey, Beatty, Kearney, & Plax (1985) found a clear multidimensional factor structure with correlations between the subscales ranging from .40 to .69. In contrast with these results, two recently developed instruments modelled directly after the PRCA24 appeared to be strictly unidimensional (Neuliep & McCroskey, 1997).

The dimensionality of the widely used Writing Apprehension Test developed by Daly and Miller (1975) is also under discussion. While the developers originally reported a unifactor structure, another study found the instrument measuring three factors that were labelled Writing Self-Concept, Affective Performance Reaction, and Evaluation Anxiety (Shaver, 1990).

In this article the dimensionality of SA and WA is studied on the basis of two relatively unknown research instruments: the Speaking Apprehension Measure (SAM) and the Writing Apprehension Measure (WAM). These instruments were explicitly developed from a multidimensional, context-based trait perspective. Compared to many other operationalisations of CA, the situation and task facets are incorporated into the instruments more systematically. The SAM measures SA in seven speaking situations and the WAM measures WA resulting from four problems a student may encounter during writing.

Another important question pertains to the presupposed discriminant validity of the SAM and the WAM. National assessment research in Dutch secondary education has shown that performance levels in Dutch language vary greatly from one school to another (Kuhlemeier & Van den Bergh, 1989; Kuhlemeier, Kremers, & Kleintjes, 1997). For speaking performance 28% of the score variance was between schools, and for composition skills it was 31% (Kuhlemeier, 1996). In general, large between-school differences suggest that educational outcomes are strongly affected by classroom and school characteristics (Scheerens & Bosker, 1997; Willms, 1992). Far fewer studies into

school effectiveness have focused on non-cognitive outcomes such as school climate and class climate (Anderson, 1982; Knuver & Brandsma, 1993; Mortimore, Sammons, Stoll, Lewis, & Ecob, 1988; Pallas, 1988). The general finding seems to be that between-school differences in the affective domain are much smaller than they are in the cognitive domain, indicating that schools are exerting more influence on students' achievements than on their attitudes and anxieties (Knuver & Brandsma, 1993; Kuhlemeier, Melse, & Van den Bergh, 1996; Pallas, 1988; Scheerens & Bosker, 1997). An important question is how much of the variation in CA is between schools and how much is between students within schools. Do all the students in some schools report high CA and all the students in other schools report low CA? Or do most schools typically have a variety of students with high and low CA? If CA is an important attribute of the school, most of the students in a given school should have similar levels of CA. That is to say, within schools, there should be few differences between students. In contrast, if CA is primarily a personal trait, there should be hardly any differences between schools in the mean level of CA. All the variance should be within schools (i.e., between students), and none of it between schools. Of course, the proportion of between-school variance will lie somewhere between .00 and 1.00, but what matters is where it lies precisely (Pallas, 1988). Trait and context-based conceptualisations of CA are both viewed as personality-specific characteristics rooted firmly in the student's personality structure. The SAM and WAM would show discriminant validity if the scores primarily reflect differences between students.

All in all, three research questions are central in this article. The first research question concerns the dimensionality of the measurement of oral and written CA and the extent to which the SAM and the WAM are trait measures or context-based measures. To what extent can the presupposed speaking situations and writing problems be differentiated empirically? Or, put in other words, do SA and WA essentially represent unidimensional or multidimensional constructs? High correlations between subscales measuring anxiety related to speaking situations or to writing tasks would support a trait-like view on SA and WA, while low to moderate correlations would suggest that SA and WA are essentially context-based constructs.

The second research question pertains to the relationship between SA and WA as subconstructs of communication apprehension. Do SA and WA represent different skill-specific constructs? Alternatively, are we essentially dealing with two equivalent forms of communication anxiety? Previously, a substantial relationship between SA and WA was expected. First, SA and WA are conceptually related. They may be conceptualised as two skill-specific forms of a more general construct, which may be labelled as language anxiety, or, communication apprehension. Second, the questionnaires are developed from the same general theoretical framework (McMillan, 1980). In both instruments, questions are posed about cognitive, affective and psycho-physiological reactions. Third, a common source underlying SA and WA may be evaluation apprehension (Geen & Bushman, 1987). SA and WA may both be elicited by the anxiety arousing belief that the teacher will evaluate one's performance and that one may possibly obtain a low grade. Fourth, SA and WA are measured with the same type of instruments (questionnaires). Multi-trait-multi-method research has shown that, in general, two constructs measured with the same type of instrument correlate higher than the same two constructs measured with different types of instruments (Campbell & Fiske, 1959).

The third research question pertains to one aspect of the presupposed discriminant validity of de SAM and WAM. How large are differences between schools in the mean

level of SA and WA? If between-school differences are very small, SA and WA may indeed be interpreted primarily as individual traits that are hardly influenced by the school which the students attend.

The answers to the research questions concerning the dimensionality of CA are not only interesting from a theoretical point of view. There might also be practical implications for the treatment of CA. If CA appeared to be a unidimensional construct and a student's level of CA was the same across situations and tasks, there would be no need to tailor the treatment to particular speaking situations or specific writing problems. If, however, CA appeared to be multidimensional, and a student might have a variety of different levels and types of CA, one type of treatment might be more effective for one student than for another one. For instance, it seems reasonable that apprehension stemming from problems in generating relevant content may require a different type of treatment than apprehension resulting from problems with the mechanics of writing or with the tailoring of the text to a particular audience.

Method

Data collection and subjects

The data on SA and WA were taken from the 1987–1988 National Assessment of Language Performance in the Netherlands (Kuhlemeier & Van den Bergh, 1989). In this study, a nationally representative description was provided of Dutch language performance, activities, attitudes and anxieties in the third year of secondary education (i.e., ninth grade). The data were collected at the end of the 1987–1988 school year. The questionnaires for speaking and writing apprehension were completed by 1448 and 1444 students, respectively. The students were drawn from 184 schools with the following breakdown for curricular track: pre-university education (19%), higher general education (19%), intermediate general education (23%), vocational technical education (21%) and vocational domestic-science education (18%); 52% were boys and 48% were girls; the mean age of the students was 15 years 6 months.

Instruments

Both questionnaires were constructed following a facet design (Guttman, 1965).

The Speaking Apprehension Measure (SAM) was developed by Marschall according to a 7*3*3 facet design (Van den Bergh, Baltzer, Marschall, Triesscheijn, & Wesdorp, 1986) and validated by Rijlaarsdam and Van den Bergh (1987). The three facets were situation, reaction type, and time. In the situation facet, seven speaking situations were distinguished: 1) reading out loud at one's own desk, 2) participating in a small-group discussion, 3) answering teacher's questions at one's own desk, 4) participating in a class discussion, 5) giving a talk in front of the class, 6) answering teacher's questions in front of the class, and 7) dramatic expression.

In accordance with the tripartite definition of attitudes (McMillan, 1980), affective, cognitive and psycho-physiological reactions were distinguished in the reaction type facet. In the time facet, anxiety before, during and after conducting the speaking task was distinguished. Crossing the three facets resulted in 63 cells and each cell was represented by one statement.

The Writing Apprehension Measure (WAM) was developed by Meuffels (1988)

according to a 4*4 facet design with two facets: task and reaction type (Van den Bergh *et al.*, 1986). In the task facet, four writing problems were distinguished: 1) general writing problems, 2) technical conventions, 3) generating relevant content, and 4) purpose and audience directness.

The first aspect of the task facet refers to general feelings of unease and fear resulting from blocks that may occur when students are asked to commit their ideas to paper. The remaining three aspects are derived from Flower and Hayes' theory (1980). The aspect of 'Technical conventions' refers to problems related to spelling, punctuation, grammar and stylistic conventions. The aspect of 'Content' pertains to anxieties that may occur when students do not have enough facts or knowledge at their disposal, or when they are not capable of generating relevant content. The latter aspect refers to problems tailoring the text to the audience and expressing one's intentions.

In the reaction type facet, a distinction is made between cognitive, affective, physiological, and behavioural reactions (McMillan, 1980). Crossing the two facets resulted in 16 cells, each cell represented by five statements.

Items were in the Likert format. Students answered the questions by indicating the extent to which they agreed with the statement or the extent to which the statement applied to them. In the analysis, negatively formulated statements were 'mirrored'. This means that the higher a scale score, the less apprehension is experienced.

Structure of the dataset

These days it is generally agreed that the appropriate technique for the statistical analysis of educational data is a multilevel one. For education is organised hierarchically, with students nested within schools. In the present dataset, scores are nested within students, and students are nested within schools. A multilevel analysis explicitly takes this nested structure into account: it yields more accurate parameter estimates and it offers extra opportunities for interpretation (cf. Goldstein, 1987).

A preliminary analysis on the SAM data showed that the three aspects of the time facet could not be distinguished empirically (Kuhlemeier, 1996). Therefore, the original 63 scores were reduced to 21 subscale scores by summing across the three statements. Eventually the SAM dataset contained 30210 subscale scores, belonging to 1448 ninth-grade students from 184 secondary schools. The reliability (Cronbach's alpha) of the 16 three-item subscales varied from .54 to .79 with a mean of .68.

For the WAM data, the original 80 scores were reduced to 16 subscale scores by summing across the five statements for each of the 16 aspects. The resulting dataset consisted of 22640 subscale scores, belonging to 1444 students from 184 schools. Reliability (Cronbach's alpha) of the 16 five-item subscales varied from .63 to .82 with a mean of .73.

For the analysis of the relation between WA and SA, 52603 subscale scores were available, belonging to 1444 students from 184 schools.

Statistical analysis

The data were analysed by means of a MultiLevel Factor Analysis (Goldstein *et al.*, 1998). This technique for the analysis of hierarchical data enables a simultaneous analysis of the factor structure at the school and student levels. Additionally it takes into account the cluster effect (Kish, 1965) and the different number of observational units at the school and student levels. Therefore it results in more correct estimates of the

standard errors (Goldstein, 1987). It would lead too far afield to describe this technique, and its application in our study, in detail. For the statistical backgrounds of MLFA we refer to Goldstein and McDonald (1988) and Longford and Muthén (1992), and for concrete applications to Longford (1990) and Raudenbush, Rowan, and Kang (1991). Here we will suffice with a short description of MLFA in relation to the research questions. For technical details about the analysis levels that are distinguished, the analysed models, the statistical significance tests, and the calculation of the reliability coefficients, the reader is referred to the Appendix.

A distinctive feature of MLFA is that the factor structure can be examined at the school and student level simultaneously. First, MLFA determines the number of factors needed to give a sufficient description of the data at each level (research question 1 and 2). In the domain of cognitive abilities, the student level structure is often more complicated than that at the school level. Thus for an adequate description of the student abilities, more factors will be needed than for a characterisation of the school's mean performance level (cf. Balke, 1991; Kuhlemeier, Van den Bergh, & Wijnstra, 1995; Kuhlemeier, 1996).

Second, MLFA provides information about the correlations between the factors at the student level and the school level (research question 1 and 2). The between-students correlations are interpreted as the (mental or attitudinal) structure of students, just like in an unilevel factor analysis. The between-school correlations represent the relationships between the factor means of the schools.

Third, in a MLFA the total factor variance can be decomposed into a between-schools component (school level), a between-students-within-schools component (student level) and a between-scores-within-student component (score level). The percentages of variance at the school and student levels reflect the extent to which an instrument measures differences between schools (research question 3) versus differences between students within schools (e.g., Balke, 1991; Kuhlemeier, 1996; Longford, 1990). In research on instructional and school effectiveness, a relatively large between-schools component suggests that educational outcomes are strongly affected by classroom and school characteristics (Scheerens & Bosker, 1997; Willms, 1992). With attitudinal data, factor variances at the student level may reflect differences between students in their individual beliefs, perceptions, and anxieties. At the school level, on the other hand, factor variances refer to differences between schools in the beliefs and feelings that are shared by all students in a school (Anderson, 1982; Raudenbush *et al.*, 1991). Analogously, if the between-students component is relatively large, SA and WA may be interpreted primarily as individual traits that are hardly influenced by the school that the students attend. The percentage of variance at the score level provides an initial estimate of the (un)reliability with which the factors at the school and student level are measured (see the Appendix).

Results

Speaking apprehension

The results of the multilevel factor analysis of SA are presented in Table 1. The upper, middle and lower sections show the analysis results at the school, student and score level, respectively. The SAM is intended to measure SA in seven speaking situations. How many factors should be distinguished in order to give an adequate description of

the SA in Dutch secondary education (research question 1)? At the student level, all seven presupposed factors are supported by the data. The student level variances – estimated as proportions of the total variances in SA – vary from .59 for 'Small group discussion' to .75 for 'Dramatic expression'. Apparently, SA differs from situation to situation.

The correlations between the students' factor means show a clear pattern. The correlations are all positive, varying from .46 to .86 with a mean of .69. In general, the seven speaking situations can be distinguished rather clearly. The correlation pattern shows some evidence for discriminant validity (Campbell & Fiske, 1959). The correlations between similar speaking situations seem to be somewhat higher than those between conceptually different situations. For instance, students do not seem to make a clear distinction between a small group discussion and a whole-class discussion (r = .86); the same goes for answering questions in front of the class versus answering questions at one's own desk (r = .86).

Table 1. Factor variance at school level (upper part), factor variances and correlations at student level (middle part; diagonal: variances; below diagonal: correlations) and residual variances at score level (lower part) for speaking apprehension

	SCHOOL LEVEL						
Speaking apprehension				.02			
	STUDENT LEVEL						
	SAM1	SAM2	SAM3	SAM4	SAM5	SAM6	SAM7
Reading out loud	.64						
Small group discussion	.53	.59					
Answering questions (seat)	.75	.71	.64				
Class discussion	.56	.86	.77	.67			
Giving a talk	.63	.63	.79	.70	.68		
Answering questions (class)	.65	.66	.86	.75	.83	.71	
Dramatic expression	.46	.58	.65	.65	.74	.72	.75
	SCORE LEVEL						
Cognitive reaction	.35	.36	.35	.27	.25	.26	.24
Affective reaction	.32	.36	.30	.31	.26	.24	.17
Psycho-physiological reaction	.35	.46	.38	.39	.42	.30	.26

At the school level, the data can be described adequately using only one factor; no distinction can be made between the seven speaking situations. This school-level factor can therefore be interpreted as the general level of SA at that particular school. The differences between schools in the level of SA and WA are small (third research question). The upper section of Table 1 shows that only 2% of the total variance in SA can be attributed to the school the student attends.

The proportions of residual variance, which cannot be accounted for by the factors at the school and student levels, provides an initial indication of the (un)reliability of the measurement (see the Appendix). The lower section of Table 1 shows that the residual variances vary from .17 to .46 with a mean of .31. Almost one-third of the total variance is therefore attributed to the score level. In other words: about two-thirds of the total variance is explained by the factors at the school and student levels. The 21 subscales

are not all equally useful for indicating the factors they are linked to. The affective reaction subscale is the best indicator of the factor of dramatic expression (the residual variance is only .17), whereas the psycho-physiological reaction subscale is a rather poor indicator of the small group discussion factor (the residual variance is no less than .46).

Writing apprehension

The results of the multilevel factor analysis of WA are presented in Table 2. The upper, middle and lower sections show the results at the school, student and score level, respectively. The WAM is intended to measure apprehension related to four writing problems. At the student level, three factors are needed to give a sufficient description of the students' WA. The student level variance, estimated as the proportion of the total variance in WA, is considerably higher for 'Generating relevant content' as for 'Technical conventions' and 'General writing apprehension' (respectively, .70, .57 and .58). Apparently, WA depends on the writing problem.

Table 2. Factor variance at school level (upper part), factor variances and correlations at student level (middle part; diagonal: variances; below diagonal: correlations) and residual variances (lower part; between brackets: residual variances for purpose- and audience-directedness) for writing apprehension

	SCHOOL LEVEL		
Writing apprehension		.08	
	STUDENT LEVEL		
	WAM1	WAM2	WAM3
General writing apprehension	.58		
Technical conventions	.85	.57	
Generating relevant content	.85	.67	.70
	SCORE LEVEL		
Cognitive reaction	.49 (.37)	.37	.20
Affective reaction	.33 (.27)	.33	.22
Physiological reaction	.32 (.34)	.30	.26
Behavioural reaction	.34 (.36)	.35	.24

The correlation between the factors for 'General writing problems' and 'Purpose- and audience-directness' did not differ significantly from 1.00. Apparently students do not discriminate between general writing problems and problems related to the purpose and the audience of the text. Consequently, in the analysis both factors were collapsed into one new factor that can be interpreted as general writing apprehension. The second factor is indicated by the four subscale scores for 'Technical conventions', and the third factor is defined by the four subscale scores for 'Generating relevant content'.

Table 2 also shows that the separability of 'Technical conventions' and 'Generating relevant content' is supported by the data (r =.67). On the other hand, students do not seem to differentiate very clearly between general writing apprehension and writing apprehension that results from problems with technical conventions (r =.85) or from problems in generating relevant content (r =.85).

At the school level, only one factor is needed to give an adequate description of the mean level of WA in the school. The between-school variance of the factor means for WA is higher (8%) as it was for SA (2%). It can be concluded that schools differ to a greater degree in the category of WA than they do in the category of SA.

The residual variances vary from .20 to .49 with a mean of .32. Just like SA, almost one-third of the total variance is not attracted by the factors at the school and student levels. The cognitive, affective, physiological and behavioural subscales indicate the factor for generating relevant content rather well. The reverse holds true for the cognitive reaction subscale as an indicator of 'General writing apprehension' (the residual variance is no less than .49).

Relationship between oral and written CA

To what extent can SA and WA be distinguished (research question 2)? Table 3 shows the correlations between the factor means of the students. At the student level, SA can be differentiated from WA clearly; the two forms of CA appear to be only weakly related. The correlations between the seven speaking situations and the three writing problems are all positive and vary from .29 to .44 with a mean of .36. Or, put in other words, measures of SA share only 13% of their variance with measures of WA.

Table 3. Correlations between speaking and writing apprehension at student level[1]

		SAM1	SAM2	SAM3	SAM4	SAM5	SAM6	SAM7
WAM1	General writing apprehension	.37	.44	.37	.39	.36	.37	.35
WAM2	Technical conventions	.42	.39	.38	.35	.35	.37	.32
WAM3	Generating relevant content	.29	.33	.30	.33	.33	.31	.38

[1] The correlation between the factor means of the 184 schools was .43 and did not differ significantly from 0.00 ($p > .05$)

At the school level, the correlation between the factor means of the 184 schools for SA and WA was .43 and did not differ significantly from 0.00 ($p > .05$). When the students in a particular school exhibit a relatively high level of SA they do not also show a relatively high level of WA.

Discussion

The first purpose of this study was to examine the dimensionality of SA and WA. The seven-situation model of SA, with each situation taking a separate factor, was fully supported by the data. Furthermore, empirical evidence was found for three out of four anxiety-provoking writing problems. The seven speaking situations and the three writing situations indeed represent separate constructs, and are not simply indications of student trait anxiety. These results suggest that the SAM and the WAM are measuring context-based CA more than trait-like CA. Oral and written CA appear to be multidimensional constructs that depend on situational and task factors. Unidimensional measures of CA may only enable a rough approximation of the student's actual

perceptions and feelings. General and simple measurements may not do justice to the fact that a student's thoughts, intentions and feelings may be ambivalent and may differ from one situation to the next or from one task to the next. Therefore, more precise and more differentiated measuring devices are needed (cf. Marsh, 1990; McMillan, 1980; Pekrun, 1988). The general recommendation would be to incorporate the situation and task facets into future instruments for measuring CA more systematically.

The second purpose of the study was to examine the relationship between SA and WA. Previously, we expected a substantial relationship between both subconstructs of general CA. It appeared that the two constructs could be differentiated clearly; correlations were rather low. Only 13% of the variance on a measure of SA can be predicted from the score on a measure of WA. This result suggests that oral and written CA may represent separate, skill-specific constructs which can hardly be considered as equivalent forms of communication apprehension.

A possible explanation for the rather weak relationship between SA and WA may be found in differences between the social contexts of oral and written communication. Spoken messages typically presuppose the presence of other persons, whereas written messages typically address an absent audience (Fondacaro & Higgins, 1985). During speaking the mere physical presence of an observing and evaluating audience may give rise to serious feelings of uncertainty and arousal (Geen, 1991; Geen & Bushman, 1987). Paivio (1965) found that students with high levels of audience anxiety were more aroused during speaking and spoke for shorter lengths of time. Because written communication does not involve face-to-face interaction, the audience may not as such constitute a strong anxiety-provoking factor. Furthermore, it is only in speaking that a student may be confronted with the self-presentational dilemma of satisfying the demands and expectations of the teacher and the peer group simultaneously (Baumeister & Hutton, 1987). A supplementary explanation is taken from distraction-conflict theory (Baron, 1986). Paying attention to others distracts subjects from task performance, thereby creating attention conflict and cognitive overload (Baron, Moore, & Sanders, 1978; Geen, 1991; Hertz- Lazarowitz, Kirkus, & Miller, 1992). In performing a difficult speaking task, one is faced with the dual problem of directing one's attention to the speaking task and the nearby audience simultaneously. When environmental demands exceed the limited information processing capacity, cognitive overload and selective attention may produce performance loss.

Another possible explanation for the rather weak relationship between SA and WA refers to differences in the extent to which linguistic problems are operationalised in the questionnaires. In the WAM, questions are focused on problems regarding the linguistic aspects of the text, whereas linguistic problems are of minor importance in the SAM. If, for instance, problems in generating relevant content were given more attention in the SAM, this would almost certainly have boosted the correlations with the WAM.

The third purpose of this study was to provide information about the differences between schools in the level of SA and WA. At the school level the measurement of SA and WA appeared to be unidimensional; only one factor was needed to provide a sufficient description of the mean level of SA or WA of the school the students attended.

Between-school differences in SA and WA were found rather small (2% and 8% of the total variance, respectively). Differences between schools were considerably smaller than they are for speaking and writing performance (Kuhlemeier, 1996). This result is in line with the more general finding that schools seem to exert more influence on the cognitive domain than on the affective domain (Knuver & Brandsma, 1993; Kuhlemeier

et al., 1996; Pallas, 1988; Scheerens & Bosker, 1997). The relatively large proportion of between-students variance indicates that SA and WA are essentially individual traits, rooted firmly in the student's personality structure. Unlike performance measures, oral and written CA are relatively unaffected by school characteristics like the composition of the student body and the effectiveness of instruction and school policies. A second explanation for the small between-school differences might be that the allocation of students to curricular tracks depends almost entirely on the student's performance; SA and WA are not used as criteria for selection. Differential selection policies do not contribute to between-school differences in SA and WA as much as they contribute to between-school differences in academic performance.

References

Allen, M., Hunter, J., & Donahue, W. (1989). Meta-analysis of self-report data on the effectiveness of public speaking anxiety reduction treatment techniques. *Communication Education*, *38*, 54–76.

Andersen, P.A., Andersen, J.F., & Garrison, J.P. (1978). Singing apprehension and talking apprehension: The development of two constructs. *Sign Language Studies*, *19*, 155–186.

Anderson, C.S. (1982). The search for school climate: A review of the research. *Review of Educational Research*, *52*, 368–420.

Balke, G. (1991, April). *Multilevel factor analysis of proficiency in English as a foreign language*. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL. [ED 334216].

Baron, R. (1986). Distraction-conflict theory: Progress and problems. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (pp. 1–40). Orlando, FL: Academic Press.

Baron, R.S., Moore, D., & Sanders, G.S. (1978). Distraction as a source of drive in social facilitation research. *Journal of Personality and Social Psychology*, *36*, 816–824.

Baumeister, R.J., & Hutton, D.G. (1987). Self-presentation theory: Self-construction and audience pleasing. In B. Mullen & G.R. Goethals (Eds.), *Theories of group behaviour* (pp. 71–87). New York: Springer-Verlag.

Bentler, P.M., & Bonet, D.G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*, 588–606.

Bergh, H. van den., Baltzer, J.E., Marschall, G., Triesscheijn, B., & Wesdorp, H. (1986). *Instrumenten ter peiling van de taalvaardigheid van leerlingen uit de derde klas voortgezet onderwijs* [Instruments for the national assessment of language performance of students in the third year of secondary education]. Amsterdam: Stichting Centrum voor Onderwijsonderzoek.

Campbell, D.T., & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81–105.

Daly, J.A., & Buss, A.H. (1984). The transitory causes of audience anxiety. In J.A. Daly & J.C. McCroskey (Eds.), *Avoiding communication: Shyness, reticence, and communication apprehension* (pp. 67–78). Beverly Hills, CA: Sage Publications.

Daly, J.A., & Miller, M.D. (1975). The empirical development of an instrument to measure writing apprehension. *Research in the Teaching of English*, *9*, 242–249.

Flower, L.S., & Hayes, J.R. (1980). The dynamics of composing: Making plans and juggling with constraints. In L.W. Gregg & E.R. Steinberg (Eds.), *Cognitive processes in writing* (pp. 31–50). Hillsdale, NJ: Lawrence Erlbaum Associates.

Fondacaro, R., & Higgins, E.T. (1985). Cognitive consequences of communication mode: A social psychological perspective. In D.R. Olsen, N. Torrance & A. Hildyard (Eds.), *Literacy, language and learning. The nature and consequences of reading and writing* (pp. 73–101). Cambridge: Cambridge University Press.

Geen, R.G. (1991). Social motivation. *Annual Review of Psychology*, *42*, 377–399.

Geen, R.G., & Bushman, B.J. (1987). Drive theory: Effects of socially engendered arousal. In B. Mullen & G.R. Goethals (Eds.), *Theories of group behaviour* (pp. 89–110). New York: Springer-Verlag.

Goldstein, H. (1987). *Multilevel models in educational and social research*. London: Charles Griffin.

Goldstein, H., & McDonald, R.P. (1988). A general model for the analysis of multilevel data. *Psychometrika*, *53*, 455–467.

Goldstein, H., Rasbash, J., Plewis, I., Draper, D., Browne, W., Yang, M., Woodhouse, G., & Healy, M. (1998). *A user's guide to MLwiN*. London: University of London, Institute of Education.

Guttman, L. (1965). A faceted definition of intelligence. *Scripta Hierosolymitana*, *XIV*, 166–181.

Hertz-Lazarowitz, R., Kirkus, V.B., & Miller, N. (1992). Implications of current research on cooperative interaction for classroom application. In R. Hertz-Lazarowitz & N. Miller (Eds.), *Interaction in cooperative groups. The theoretical anatomy of group learning* (pp. 253–280). New York: Cambridge University Press.

Kish, L. (1965). *Survey sampling*. New York: Wiley.

Klopf, D., & Cambra, R. (1979). Apprehension about writing and speaking. *Psychological Reports*, *45*, 530.

Knuver, A.W.M., & Brandsma, H.P. (1993). Cognitive and affective outcomes in school effectiveness research. *School Effectiveness and School Improvement*, *4*, 189–204.

Kuhlemeier, H. (1996). *Taalvaardigheid, taalactiviteiten en taalattitudes: een validatiestudie* [Language proficiency, language activities and language attitudes. A validation study). Dissertation. Arnhem, Netherlands: National Institute for Educational Measurement.

Kuhlemeier, H., & Bergh, H. van den (1989). *De proefpeiling Nederlands: een onderzoek naar de haalbaarheid van peilingsonderzoek in het voortgezet onderwijs* [National assessment of language performance: A feasibility study in secondary education]. Arnhem, Netherlands: National Institute for Educational Measurement.

Kuhlemeier, H., Bergh, H. van den, & Wijnstra, J. (1995, April). *Multilevel factor analysis applied to national assessment data*. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA. [ED 388683].

Kuhlemeier, J.B., Kremers, E.J.J., & Kleintjes, F. (1997). *Gebruik, betrouwbaarheid en moeilijkheid van de eerste generatie afsluitingstoetsen basisvorming schooljaar 1994/95* [Use, reliability and difficulty of the first generation final tests for basic education schoolyear 1994/95]. Arnhem, Netherlands: National Institute for Educational Measurement.

Kuhlemeier, H., Melse, L., & Bergh, H. van den (1996). Comparison of two German language courses in Dutch secondary education. *Studies in Educational Evaluation*, *22*, 181–205.

Longford, N.T. (1990). Multivariate variance component analysis: An application in test development. *Journal of Educational Statistics*, *15*, 91–112.

Longford, N.T., & Muthén, B.O. (1992). Factor analysis for clustered observations. *Psychometrika*, *57*, 581–597.

Magnussen, D. (1981). Problems in environmental analysis – An introduction. In D. Magnussen (Ed.), *Toward a psychology of situations: An international perspective* (pp. 3–9). Hillsdale, NJ: Lawrence Erlbaum.

Marsh, H.W. (1990). The structure of academic self-concept: The Marsh/Shavelson model. *Journal of Educational Psychology*, *82*, 623–636.

McCroskey, J.C. (1970). Measures of communication-bound anxiety. *Speech Monographs*, *37*, 269–277.

McCroskey, J.C. (1977). Oral communication apprehension: A summary of recent theory and research. *Human Communication Research*, *4*, 78–88.

McCroskey, J.C. (1984). The communication perspective. In J.A. Daly & J.C. McCroskey (Eds.), *Avoiding communication: Shyness, reticence, and communication apprehension* (pp. 13–38). Beverly Hills, CA: Sage Publications.

McCroskey, J.C., Daly, J.A., & Sorensen, G. (1976). Personality correlates of communication

apprehension: A research note. *Human Communication Research*, *2*, 376–380.
McCroskey, J.C., Richmond, V.P., & Davis, L.M. (1986). Apprehension about communication with supervisors: A test of a theoretical relationship between types of communication apprehension. *Western Journal of Speech Communication*, *50* (Spring), 171–182.
McCroskey, J.C., Beatty, M.J., Kearney, P., & Plax, T.G. (1985). The content validity of the PRCA-24 as a measure of communication apprehension across communication contexts. *Communication Quarterly*, *33*, 165–173.
McMillan, J.H. (1980). Attitude development and measurement. In J.H. McMillan (Ed.), *The social psychology of school learning* (pp. 215–245). New York: Academic Press.
Meuffels, B. (1988). Schrijfangst-Faalangst [Writing apprehension and fear of failure]. *Tijdschrift voor Taalbeheersing*, *10*, 28–46.
Mortimore, P., Sammons, P., Stoll, L., Lewis, D., & Ecob, R. (1988). *School matters*. Berkeley, CA: University of California Press.
Neuliep, J.W., & McCroskey, J.C. (1997). The development of intercultural and interethnic communication apprehension scales. *Communication Research Reports*, *14*, 385–398.
Paivio, A. (1965). Personality and audience influence. In B. Maher (Ed.), *Progress in experimental personality research* (pp. 127–173). New York: Academic Press.
Pallas, A.M. (1988), School climate in American high schools. *Teachers College Record*, *89*, 541–554.
Pekrun, R. (1988). Anxiety and motivation in achievement settings: Towards a systems-theoretical approach. *International Journal of Educational Research*, *12*, 307–323.
Raudenbush, S.W., Rowan, B., & Kang, S.J. (1991). A multilevel, multivariate model for studying school climate with estimation via the EM algorithm and application to US high-school data. *Journal of Educational Statistics*, *16*, 295–330.
Richmond, V.P., & McCroskey, J.C. (1998). *Communication apprehension, avoidance, and effectiveness* (5th ed.). Boston: Allyn & Bacon.
Rijlaarsdam, G., & Bergh, H. van den (1987, March–April). *The development and validation of the Speaking Apprehension Measure*. Paper presented at the Annual Meeting of the International Oracy Convention, Norwich, England. [ED 285236].
Scheerens, J., & Bosker, R.J. (1997). *The foundations of educational effectiveness*. Amsterdam: Pergamon.
Shaver, J.P. (1990) Reliability and validity of measures of attitudes toward writing and toward writing with the computer, *Written Communication*, 7, 375–392.
Spielberger, C.D. (1966). Theory and research on anxiety. In C.D. Spielberger (Ed.), *Anxiety and behavior* (pp. 3–20). New York: Academic Press.
Willms, J.D. (1992). *Monitoring school performance. A guide for educators*. London: Falmer Press.
Witworth, R.H., & Cochran, C. (1996). Evaluation of integrated versus unitary treatments for reducing public speaking anxiety. *Communication Education*, *45*, 306–314.

Received 28 July 1998; revised version received 22 May 2002

Appendix: Analysis levels, statistical models and significance tests

Analysis levels

In line with the structure of the dataset – subscale scores nested within students which themselves were, in turn, nested within schools – three levels were distinguished: school (level 3), student (level 2) and score (level 1). Factor variation is decomposed into three parts: 1) a between-schools component (school level), 2) a between-students-within-school component (student level), and 3) a between-scores-within-student component (score level).

In this variance-component model, the variances and covariances of the factors are modelled at the student and school levels. The residual variance, which cannot be accounted for by these factors, is represented at the score level (Raudenbush *et al.*, 1991). The first level is the measurement level; it describes the relationship between the observed scores and the factors in the structural part of the model, the between-school and within-school level (level 2 and 3). At level 1, the intra-student-between-scores level, a separate residual variance is estimated for each subscale. The degree to which the subscale scores indicate the factors can, therefore, differ from one subscale to the next. The proportion of residual variance at the first level provides insight into the 'uniqueness' of the factors at the second and third levels. In other words, the residual variance represents that part of the total variance which cannot be attributed to these factors and which, given the model, can be interpreted as error variance. Hence the factors at the within-school and between-school levels are corrected for attenuation (Raudenbush *et al.*, 1991).

Statistical models

In the analysis we distinguish two models which only differ from each other in the factor structure at the second and third levels; the structure at the first level is identical each time. At this 'between-scores-within-students level' we estimate a separate residual variance for each subscale. The matrix of residual variances at the score level therefore is a diagonal matrix with (5*2=) ten elements. The extent to which the subscale scores indicate their factor may differ per subscale. The residual variance represents the part of the total variance which cannot be ascribed to the apprehension factors. Given the model, it represents the instability of the subscale scores within a student. According to traditional reliability theory it can be interpreted as 'noise'. This way the factors at the within- and between-schools levels have been corrected for unreliability (cf. Raudenbush *et al.*, 1991).

Basic model (Model I)

Initially, the structure of the SAM data was described using a seven-situation model (Model I). In this basic model, there is a separate factor (both at the school and the student levels) for each situation. Each situation factor has three indicators: the subscale scores for the cognitive, affective and psycho-physiological reactions. The covariance matrix at the student level is a 7*7 symmetrical matrix with 28 elements, 7 of which are variances and 21 are covariances. The between-schools covariance matrix is also a 7*7 symmetrical matrix with 28 elements.

Analogously, for the WAM data a separate factor is posited for each of the four problem types of the task facet. Each factor is indicated by the subscales for the cognitive, affective, behavioural and physiological reaction. At the school and student levels a 4*4 covariance matrix is estimated.

Simplified model (Model II)

Inspection of the covariance matrices of the basic model provided a first impression of the dimensionality of the data. Subsequently, an attempt was made to simplify the model without losing information. Two criteria were used. In order to maintain a factor as an independent entity within the model, its variance had to be significantly greater than zero ($p < .05$); if this was not the case, the given factor was removed from that specific level of the model. Secondly, the correlation with all other factors had to be significantly smaller than one (or greater than -1); perfectly

correlating factors were collapsed into a single factor. In the specific case at hand, the result was a drastically simplified model (Model II).

Significance tests

The multilevel analysis was conducted with the MLwiN program (Goldstein *et al.*, 1998) using the IGLS estimation method (Iterative Generalised Least Squares). The significance tests of the differences between variances and covariances (correlations) were conducted using a multiple comparison procedure (Goldstein, 1987, p. 29). If HO is true, this procedure results in a statistic which is approximately chi-square distributed. In statistical tests, the 5% significance criterion was applied.

The relative fit of the simplified model (Model II) as opposed to the basic model (Model I) was evaluated by the difference in -2 Log Likelihood between both models. This difference is asymptotically chi-square distributed with the corresponding difference in the number of degrees of freedom (Bentler & Bonet, 1980). On the basis of this test, Model I was rejected in favour of Model II. It is Model II that is reported in the results section.