

IEA 1958-2008: 50 Years of Experiences and Memories

EDITED BY:

Constantinos Papanastasiou

Tjeerd Plomp

Elena C. Papanastasiou

Volume 1



Cultural Center of the Kykkos Monastery

Nicosia 2011

ISBN 978-9963-645-77-0

CHAPTER 5

The International Study of Writing

Saali Takala

Introduction

I have been asked to write an account of the IEA International Study of Writing for the present volume. The untimely passing away of Alan C. Purves, who took the initiative to propose and plan the study, and of Anneli Vähäpassi, who chaired the International Steering Committee, meant that I stood next in line as I was the former International Project Coordinator. My chapter will basically be a narrative of the main activities, but will also provide some glimpses of events and processes which are not reported in the two published books on the subject (edited by Gorman, Purves & Degenhart, 1988; and by Purves, 1992). In addition, I will give some examples of the dissemination of information about the project and of the impact it seems to have had. This will be a personal account and it will be a selective one, as after almost 30 years, I had access to only a small part of relevant sources. Memory is also selective and this became very clear as, browsing the documents I did have access to, I began to recall events I had forgotten and I relived emotions I had felt (cf. Takala 1985/1986). For a perceptive discussion of the IEA written composition study, see also Hartmann (1995) and de Glopper (1995).

I will begin with a brief look at the activities of IEA in the area of humanities prior to the writing study. I recall that this prior history occasionally emerged in discussions within the writing project. It is also very clear that our study was planned in close adherence to the 'IEA model' in comparative studies of educational achievement, and so I will discuss the design and the conduct of the study, focusing on the various conceptual models we developed for the study. These are prob-

ably the study's most significant contribution to language education. I then discuss funding at some length, as problems related to resources led to delays, and a variety of problems and tensions emerged requiring many ad-hoc solutions. The main findings are briefly summarized and the chapter concludes with some reflections on the endeavor.

Context of the Study

During the Six Subject Survey, IEA conducted pioneering studies of English and French as a foreign language. I was involved in the Six Subject Survey in Finland in several capacities: as a translator of many instruments; as a member of the national team for English, writing some items that were ultimately included in the English tests; as a tester who administered tests in some schools, carrying with me a tape recorder as not all schools had one to deliver the test of listening comprehension; and finally, as a co-author of the national report. The studies coordinated by John B. Carroll for French (1975) and Glyn Lewis and Carolyn Massad for English (1975) provided a valuable model on how to conduct a study of language achievement and how to relate the results to a number of contextual and individual variables. It is noteworthy that even writing and speaking were tested—as even today, a test of speaking is often excluded from large-scale surveys.

More immediately relevant for the writing project were the study of reading comprehension in the mother tongue (Thornbike, 1973) and the study of literature (Purves, 1973). The study of literature, in particular, was of interest as it highlighted the cultural embeddedness of literature instruction and of literary appreciation. Alan Purves had continued the exploration of the data and published a study of reading and literature education (Purves, 1979) in New Zealand.

An extensive plan to initiate a project on "Language of instruction and school achievement of linguistic minorities" was proposed by Dr Stacy Churchill, OISE, at the Enschede General Assembly in August 1983. The plan was not approved. In retrospect, this is regrettable as the

study would have anticipated very well the current, and increasing, concern with the topic. Another partly missed opportunity in the area of language education was the Language Education Study (LES, 1993-1996), which managed to implement only the first phase of the planned three phases. Phase 1 gathered information on language education at the national level (sociolinguistic context, language policy, language curriculum and assessment, language teaching, and professional support), school level (characteristics of schools and language teachers, provision and organization of language learning in schools, organization of curriculum, and assessment in the classroom), and student level (level of proficiency, attitudes, and aspirations). The interesting results were reported in a survey edited by Dickson and Cumming (1996). It is only now, almost 20 years later, that an international survey of achievement in foreign languages is being carried out as an EU project-- European Indicator of Language Competence (EILC). Information on this project is available at: http://europa.eu/legislation_summaries/education_training_youth/lifelong_learning/c11083_en.htm

My background is second/foreign language education, which has a very long history of international interaction and exchange. Since the 1960s, applied linguistics has provided many new insights but it also has tended to be dominated by experts in foreign language (L2) education. While the L2 profession has thus been fundamentally internationally minded, at the time when the writing study was launched, it seemed that mother tongue education (L1) was more closely linked to the agenda of nation building and/or the construction and maintenance of national identity and values. It would appear that in comparison to L2 education, mother tongue educators had relatively few international contacts. In fact, the International Mother Tongue Education Network (IMEN) was not set up until 1984. A good survey of its work can be found in Herrltz, Ongstad & van den Ven (2007).

There were some exceptions to this pattern. In the summer of 1966 an Anglo-American seminar was arranged on the teaching of English.

Substantial differences were discovered between the attitudes towards English of teachers in England and America. Squire and Applebee (1969) published a survey of teaching of English in the United Kingdom and made a number of comparisons to their 'native' American practice. Note, however, that this rare instance of international comparison dealt with the same language, English, and was thus limited to a section of the Anglophone community.

This, in a nutshell, was the situation in the late 1970s when Alan Purves, who had a unique knowledge of the international context due to his many contacts and networks, started planning a truly international new project in the domain of the mother tongue. The fact that he chose to focus on writing was probably related to the fact that there had been a strong revival of interest in the assessment/testing of free writing to supplement - or replace - the indirect, 'objective' testing of writing, which had emerged especially in the United States as a response to the allegedly disastrously low reliability in rating free writing. Writing had also become an important area for explorations in cognitive psychology. In retrospect, it seems to me that there had been sufficient progress made both in the theory and practice of 'free' writing and in its assessment, and as a result the time was ripe for a comparative study of writing (e.g., Petrovsky & Bartholomae, 1986).

Project Approach

Alan Purves formed an international team that drafted a proposal to be submitted to the General Assembly, held at the University of Jyväskylä, Finland in August 1980. The other team members consisted of Dr Eva Baker from UCLA, Mrs Anneli Vähäpassi from the University of Jyväskylä and Dr Hildo Westdorp from SCO in the Netherlands. Eva Baker and her colleagues at the Center for the Study of Evaluation at UCLA had carried out extensive work on the assessment of writing (e.g., Smith, 1978); Hildo Westdorp was a recognized expert in mother tongue education in the Netherlands, who had been

working on a thorough review of methods used in the assessment of writing (Westdorp, 1978); and Anneli Vähäpassi was a leading expert in Finland who had extensive experience conducting national assessments in the mother tongue.

Their proposal was approved. There was a growing awareness of the importance of literacy, and there was increasing worry over the alleged deterioration of students' writing ability--frequently seen as an indication of inadequacies in the school systems. Introducing students to written language (reading and writing), and thus promoting literacy, had traditionally been seen as one of the principal tasks of the school. There was also a discernible trend in that some countries (e.g., the UK, Australia, the Netherlands, Sweden/Lindell 1980, Finland) had begun or were beginning to assess systematically the effectiveness of teaching and learning writing. However, most of the countries that ultimately participated in the study had never carried out a large-scale empirical survey of writing in their school systems.

Initially following the traditional IEA approach, the main idea was to compare the level of achievement in writing; as the project developed, however, doubts soon emerged about the feasibility of strict comparisons between countries on an indicator for achievement. This meant that ultimately the study paid increasing attention to contextual and cultural variation. This was, indeed, one of the research tasks from the very beginning, but a strong hope for reliable comparisons lingered in some participating countries. The study was accordingly designed to accomplish the following research tasks:

- to contribute to the conceptualization of the domain of writing, and particularly the domain of school-based written composition.
- to develop an internationally appropriate set of writing tasks and a system for assessing composition that would be applicable across countries, school systems and languages.
- to describe recent developments and the current state of

instruction in written composition in the participating countries/school systems (using very extensive curriculum and teachers questionnaires).

- to identify factors that would explain differences and patterns in the performance of written composition and other outcomes, with particular attention to cultural background, curriculum, and teaching practices.

The study accordingly examined teaching and learning of written composition in the schools in order to identify the beliefs and conventions associated with written composition. The study also endeavored to find factors explaining differences and patterns in the performance of written composition and other outcomes, with particular attention on cultural background, curriculum, and teaching practices. Very extensive national, school, teacher and student questionnaires were employed to collect data for such explanatory analyses. Six types of writing were assessed (reflective, personal, philosophic, argumentative, persuasive, and literary) on four dimensions: style and tone, overall impression, content, and organization. The data were collected in 1985.

The design presented in Figure 1 shows that the writing study was planned using the traditional model employed in earlier IEA studies. Participating countries included: Chile, England, Finland, Hamburg-Germany (FRG), Hungary, Indonesia, Italy, the Netherlands, New Zealand, Nigeria, Sweden, Thailand, the United States and Wales. The International Coordinating Center was initially located at the University of Illinois at Urbana-Champaign (UIUC).

The study included three populations: students near the end of primary schooling (Pop. A); students near the end of compulsory schooling (Pop. B); and students near the end of academic secondary schooling (Pop. C).

Participating countries selected writing samples, translated them into English on the basis of instructions provided, and provided preliminary ratings. Two international scoring sessions were arranged to rate

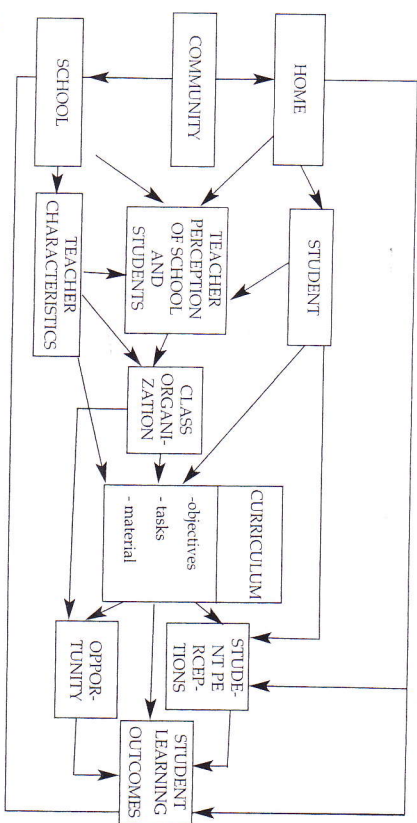


Figure 1. Design of the Writing Study
(Gorman, Purves & Degenhart, 1988, p.10)

the scripts in teams. This provided feedback so countries could see how closely the international juries agreed with their ratings. This procedure also produced benchmarks that could be used as supports in national rating sessions.

Development of Conceptual Models for the Assessment of Writing: the Main Lasting Contribution of the Project?

It seems to me that the main long-term contribution of the written composition study differs from a 'typical' IEA study. This study did not lead to sufficiently reliable and valid comparisons of the level of achievement among the participating countries, or to clear patterns that would explain the achievements. I believe that its main contribution was the development of quite novel and useful approaches to the assessment of writing (the first three research tasks), thus yielding a better conceptualization of the domain. The project was not only a huge challenge, but it was also an opportunity for raising awareness about the complexities of such an endeavor and about the strong influence of cultural and contextual factors in the teaching and assessment of writing.

While it is true that there was very useful literature to draw on (Cooper & Odell, 1977; Diederich, French & Carlton, 1961), the approaches used tended to be quite narrow in focus, building more on tradition than exploring new avenues. Our models for the domain of writing and for the assessment of writing were favourably received when the study was presented at several international conferences, including AERA (American Educational Research Association) and TESOL (Teachers of English to Speakers of Other Languages). I recall that one eminent reviewer referred to our work in print as "remarkable". I will illustrate some of the models in the following section.

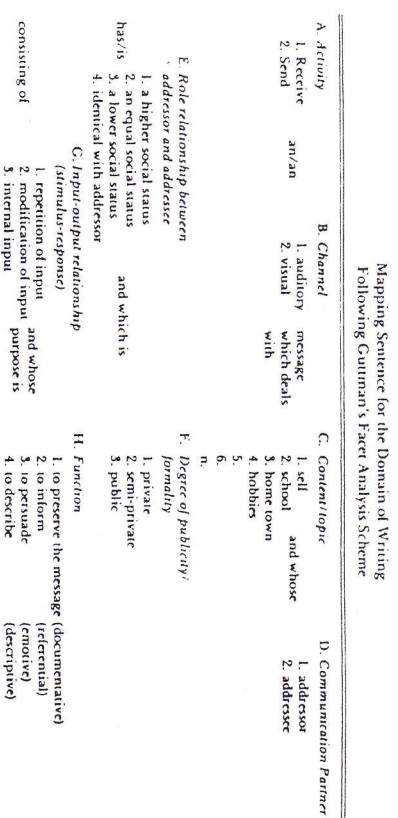
During the first months after the launch of the project, there was very intensive work devoted to developing models to get a better grasp of the construct of writing. A good example is the domain of writing developed by Vähäpassi and presented in Figure 2. It has become quite well known in in the mother tongue education profession.

Dominant Intention/ Purpose	Cognitive Processing		Reproduce	Organize/Reorganize	Invent/Generate
	Primary Audience	Content			
To learn (metalingual)	Self	Facts	Ideas	Visual images, facts, mental states, ideas	Ideas, mental states, alternative words
To convey emotions, feelings (emotive)	Self, others	stream of consciousness	copying, taking dictation	personal diary, personal letter, portrait	reflexive writing, personal essays
To inform (referential)	Others	quote	fill in form	narrative report, news, fiction, telegram, announcement, circular	expository writing, didactic, technical, description, biography, book review, circular
To convince/ persuade (conative)	Others	citation from authority/expert	statement of opinion	letter of application, advertisement, letter of advice	argumentative/ persuasive writing, critical essay/article
To entertain, delight, please (poetic)	Others	quote poetry and prose	create a story	word portrait or sketch	entertainment writing, parody, rhymes
To keep in touch (phatic)	Others	greeting card	postcard	personal letter	humorous greeting

Figure 2. Definition of the domain of writing in the Writing Study (Gorman, Purves & Degenhart, 1988, p. 22).

Reference is made to it, for instance, in the key work by Weigle (2002). The model was the first to provide a systematic classification of writing which unites the cognitive and functional dimensions of writing. It indicates how the various functions (purposes) of writing combine with different levels of cognitive processing to produce different types of written products (exemplified in the cells). It also takes note of the audiences of writing and of the primary content of writing. The audiences can range from the self and familiar recipients to an unknown general audience. One of the most important content variation dimensions is whether the content has a temporal, spatial or notional/conceptual foundation (leading to narrative, descriptive, or expository/argumentative writing, respectively).

The domain definition provided a very useful conceptual basis for discussing the sampling of writing tasks (nine tasks were developed, with one task serving as an anchor for the three populations; see



Different configurations of variables lead to different theoretical modes (narrative, exposition, argumentation, etc.)

Examples: A2 + B2 + C2 + D2 + E1 + F3 + G2 + H1 = a personal letter to a friend

A2 + B2 + C2 + D1 + E3 + F2 + G4 + H1 = a letter of application

Figure 3. Mapping sentences for the domain of writing (Purves, Söter, Takala & Vähäpassi, 1984, p. 392).

Figure 4 below). It also facilitated the development of a scoring scheme.

In the attempt to get a better grasp of the domain we also tried out Luis Guttman's approach to domain definition (e.g., Guttman, 1970), the facet analysis.

The approach proved interesting and indicated how vast the variety of written products could be. The model is not included in the published reports but appeared in the journal, *Research on the Teaching of English* (Purves, Söter, Takala & Vähäpassi, 1984).

In addition to the specification of the domain of writing, it was felt important to develop a model which would provide a conceptual basis for a rating scheme. There was no shortage of rating systems, beginning with the pioneering work by Diederich et al. (1961). Work done on "primary trait" scoring (e.g., Lloyd-Jones, 1977) was also a potential approach but it was not adopted as such. In primary trait scoring performances are evaluated by limiting attention to a single criterion (or a few selected criteria). This criterion or these criteria are based upon the trait or traits determined to be essential for the successful performance of a given task. For example, persuasiveness is the primary trait in an argumentative task. As far as I can recall, it was felt that it was too 'rigid' in demanding a closely genre-tied rating, whereas it was possible that not all cultures shared the same view of the appropriate realizations of different genres. Kaplan's early work on cross-cultural rhetoric (1966), as well as his subsequent work, also suggested the advisability of some caution in this respect.

Figure 4 presents the distribution of the nine different kinds of writing tasks across populations. The tasks will not be discussed in further detail. Suffice it to say that task 9, a letter of advice to a younger fellow student on how one should write in order to get good grades in the school, had a double function: it tapped the instructive text type but it also provided also an opportunity to analyze what views students had of what counts in writing at school. It was initially labeled "bizarre" by

Dominant intention/ Purpose	Primary Cognitive Demand		
	I Reproduce	II Organize Reorganize	III Invent Generate
1. To learn (metalingual, mathetic)		Summary (B,C) Paraphrasing (A)	
2. To convey emotions, feelings (emotive)		Narrative/Personal story (A,B)	Open essay (B,C)
3. To inform (referential)	Letter to an uncle describing a bicycle (A,B) Self description in letter to pen-pal (A,B) Formal note to head of school (A,B) Message to family (A) Application letter (B,C)	Letter of advice to younger student (B,C) Describing an object (A,B) Describing a process (B,C)	Reflective essay (B,C)
4. To convince/ persuade (conative)		Application letter (B,C) Letter of advice to younger student (B,C)	Persuasive/ argumentative essay (A,B,C)
5. To entertain, delight, please (poetic)			Open essay (B,C)

Figure 4. Distribution of tasks across Populations A, B and C (Gorman, Purves & Degenhart, 1988, p. 33)

one of the participants but turned out to be a very good idea. For a variety of reasons, it held a great appeal to students.

Instead, early work done by Canale and Swain (1980) on modelling communicative language ability, and by Lyle Bachman, who at that time was working on pioneering projects at UIUC (eg., Bachman &

Palmer, 1982), appeared promising. Combining these with the latent trait approach led to the development of the model for student writing represented in Figure 5. It breaks down the general construct of student writing into writing competence and writing preferences, and elaborates writing competence into the two latent traits of discourse-structuring and text-producing competence, and continuing such an elaboration to end up with manifest variables (rated scores).

The above description of model-building is not comprehensive. Similar work was done, for instance, on outlining different schemes of linking raters with each other to make it possible to estimate inter-rater agreement.

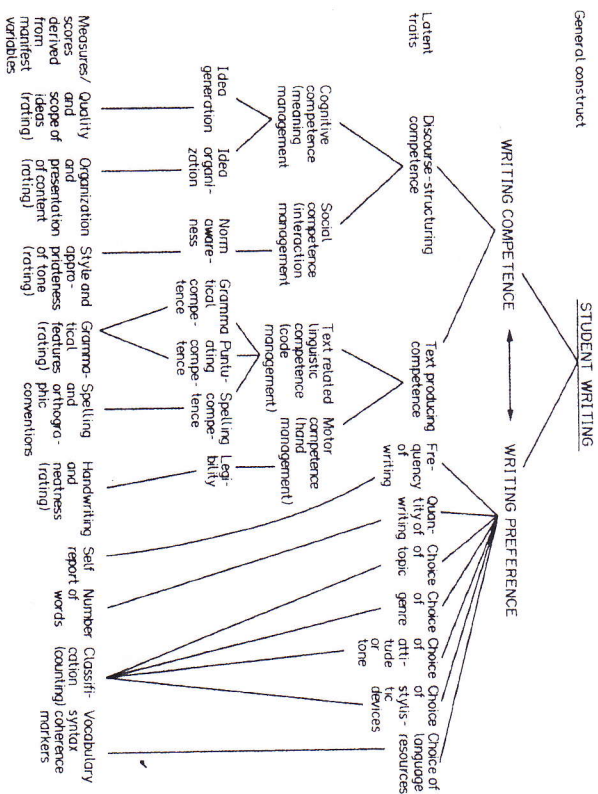


Figure 5. Model used to develop the general rating scheme (Gorman, Purves & Degenhart, 1988, p. 43).

Problems due to Inadequate Resources and Solutions Attempted

The Spencer Foundation had initially provided funding to get the project off the ground (1981), but there was no success in raising further funding in the spring and summer of 1982. After a period of great anxiety, the Spencer Foundation extended its support by about two more years. Such support made it possible to hire a project coordinator (myself) at UIUC, an assistant/secretary, and later on part-time psychometric assistance. However, one of my most vivid recollections of the project is the constant worry over funding.

The project funding was running out again in the autumn of 1984 and I had to return to my home institute in Finland. I brought with me a massive dataset that needed to be analyzed to provide an international backdrop to national analyses. In the end, however, many countries had to write these reports without the international analyses. Understandably, this led to growing dissatisfaction in those countries that had a definite deadline for completing a national report.

The Institute for Educational Research at the University of Jyväskylä, Finland, was very helpful in providing data analysis services, initially under a brief contract with the IEA. However, this was a strain, as the data analysis resources were limited and there were numerous other projects to serve. Some time later, Elaine Degenhart arrived from Illinois and took on the main responsibility for practical coordination for the next three years. Although she was funded, this lasted for a very limited amount of time. Thus the situation was such that the coordinating center was now located in Finland but it had practically no resources allocated for the work. The participating countries, with a strictly pre-determined schedule and budgets, were clamouring for the international reports as a basis for their national reports. An emergency meeting was convened in Hamburg in early February in 1985 to try to find ways of making faster progress. Concrete plans were suggested and they were helpful, but as no major new funding was forth-

coming, faster progress was clearly not possible.

This called for creative fundraising: I applied for a vacant professorship in language didactics at the University of Helsinki and as I seemed to be the most qualified among the applicants I was asked to act as a *pro tempore* professor for a whole academic year. This released my salary and the Institute invested it in the IEA study. I subsequently also acted for a term as a substitute professor of applied linguistics at my own university and the released money was again used for the IEA study. We also managed to obtain a 3-month Fulbright scholarship for Alan Purves and Elaine Degenhart to come and work at the Institute in the summer of 1989. There was no summer holiday (a sacred thing in Finland and much needed after the hard winters!) during that summer; instead, there was intensive work even on weekends, focused on data analysis and writing up draft texts.

My home institute was by no means the only party to assist the project to push on despite the constant funding problems. For the project, vitally important was the support that members of the extended International Steering Committee managed to arrange in hosting meetings: Judith Kadame-Fülöp from the Hungarian National Institute for Education (OPJ) hosted a meeting in Eszgerom and Budapest, Thomas Gorman from the National Foundation for Educational Research in England and Wales (NFER) in Slough, Pietro Lucisano at the European Centre for Education (CEDE) in Frascati, Hildo Wessdorp from SCO in Amsterdam, Eva Baker at UCLA and Anneli Vähäpassi in Jyväskylä.

The chronic lack of funds created a number of problems:

- The international results were delayed and many national reports had to be written without having sufficient international comparative analyses to support their work (e.g., de Glopper, 1985; IEA IPS, 1988; Gubb, Gorman & Price, 1987; Lamb, 1987; Löfqvist, 1988).

- It was not possible to secure top statistical advice and services as the funding base was so insecure.

- The delays inevitably led to some tensions, occasionally quite serious, among the participants. Unreasonable amounts of work were required of many people under unreasonable deadlines.
- The use of partial least squares modelling (eg., Sellin, 1995) as one of main statistical methods for multivariate analyses was an extra challenge as it was less known in Finland than in Germany, requiring intensive correspondence.

The pressures we felt came also from inside the IEA itself. If my memory serves me, the reporting of the math study was much delayed and I recall that the vocal criticism for the delay caused pressures for the writing study to 'deliver' quickly, beyond what was at all reasonable given the problems in staffing due to the lack of funds. I remember objecting strongly to speeded-up reporting as this would have led to problems with quality.

The contributions of all the parties mentioned above made it possible to report the main outcomes of the study in two volumes (Gorman, Purves & Degenhart, 1988; Purves, 1992). As for the main findings, the following summary on the IEA website offers an apt and succinct account:

1. The construct "written composition" was found to be sited in a cultural context and so could not be considered a general cognitive capacity or activity. Marked variation across the countries existed in both ideology of the teachers and in instructional practices. Written performance was also found to be task dependent.
2. Good compositions from different countries shared common qualities related to handling of content and appropriateness of style, but these qualities had their national or local characteristics in organization, use of detail, and other aspects of rhetoric.
3. Students across educational systems had in common a sense of the importance of the written product and its surface features. Beneath that commonality, however, there was national variation in the perception of what is valued.
4. In most of the countries, girls were treated differently than boys in the

provision of writing instruction and in the rating of writing performance, particularly at the primary and lower secondary school levels where women largely provided instruction. In such a milieu, the most successful students were girls, and gender itself, or gender in combination with certain home variables, was the most powerful predictor of successful performance, particularly on the more 'academic' tasks.

5. Differences between the ratings of student writing were not explained by differences in instruction. They were, however, accounted for by factors involving the characteristics of the home, the reinforcement provided by parents, and the cultural values of the community.

http://www.iea.nl/written_composition.html (accessed July 18 2011)

Impact of the Study

I have already indicated that, for most of the participating countries, the written composition study represented a new venture. As in the long history of the IEA's work, participation in its projects offered hands-on training in how large-scale assessments of writing could be carried out. I will give a couple of specific examples I am familiar with.

In Italy, the very extensive curriculum and teacher questionnaires were used as a basis for organizing in-service training for teachers. These tools were felt to be very useful for raising awareness of the range of options in teaching and assessing writing. In Finland, the Finnish national coordinator, Anneli Vähäpassi, invited all teacher education departments in Finland to send representatives to the national ratings sessions; several of these participated in writing chapters in Finnish national reports. She also developed booklets drawing on the IEA tasks and their rating schemes, and she made these available to schools/teachers on a true-cost basis.

The study brought together a large number of mother tongue specialists from all over the world and helped to establish much-needed international networks. For me personally, the opportunity to work closely with such an eminent scholar as Alan Purves was a source of

constant inspiration leading to the most productive phase in my career. In addition to our joint efforts for the project, we often discussed the nature of international comparative research. On the request of the IEA Chairman, Neville Postlethwaite, we conducted a survey of what the participating countries saw as the main advantages of conducting cross-national surveys of educational achievement. Alan referred to this in an article (1987) in the *Comparative Education Review* (included in this book).

The initiative to establish an international essay database that would allow secondary analyses might have had a potentially great impact had it been successful. Some countries sent their scripts to the putative centre, but again, funding did not materialize and no such 'clearing-house' was established. One national coordinator noted that the scripts sent in simply vanished, never to be seen again.

The study gave rise to a number of conference presentations and publications over and above the international reports. Articles were published in a number of prominent journals, including the *Comparative Education Review* (Takala & Vähäpassi, 1987), *Evaluation in Education* (Purves & Takala, 1982), *Written Communication Annual* (edited by Purves, 1988).

As noted above, the study was accompanied by considerable pressure and anxiety over a number of years, and a frightening personal experience of lack of sleep over several days due to a very bad jet lag. However, these were more than compensated by becoming acquainted and working with great colleagues all over the world. Life with the study was sometimes quite exhausting, occasionally quite frustrating, but always exciting, interesting and instructive.

References

- Bachman, L. F., & Palmer, A. S. (1982). The construct validation of some components of communicative proficiency. *TESOL Quarterly* 16 (4), 449-465.

- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics* 1 (1), 1-47.
- Carroll, J. B. (1975). *The teaching of French as a foreign language in eight countries*. Stockholm: Almqvist & Wiksell.
- Cooper, R. C., & Odell, L. (Eds.). (1977). *Evaluating writing: describing, measuring, judging*. Urbana, IL: National Council for Teachers of English.
- de Glopper, K. (1985). *Schrijprestaties in het voorgezet onderwijs. Een onderzoek naar de stevigheid van leerlingen in het der der de leerjaar LTO, LNHO, MAVO, HAVO en VWO* [Writing performance in secondary schools]. Stichting Centrum voor Onderwijsonderzoek van het Universiteit van Amsterdam. Den Haag: SVO.
- de Glopper, K. (1995). Instruction and assessment in written composition in primary and secondary schools. *European Education* 27 (1), 71-86.
- Degenhart, E. & Takala, S. (1988). Developing a rating method for stylistic preference. A cross-cultural pilot study. In A.C. Purves (Ed.), *Writing across cultures. Issues in contrastive rhetoric*, (pp. 79-106). Newbury Park: Sage.
- Dickson, P. & Cumming, A. (Eds.). (1996). *Profiles of Language Education in 25 Countries*. UK: National Foundation for Educational Research.
- Diederich, P. B., French, J. W., & Carlton, S. T. (1961). *Factors in judgments of writing ability. Research Bulletin* 15. Princeton, NJ: Educational Testing Service.
- Gorman, T. P., Purves, A. C., & Degenhart, R. E. (1988). *The IEA Study of Written Composition I: The International Writing Tasks and Scoring Scales*. Oxford: Pergamon Press.
- Gubb, J., Gorman, T., & Price, E. (1987). *The study of written composition in England and Wales*. Windsor: NFER-Nelson.
- Gutman, L. (1970). Integration of test design and analysis. In "Toward a theory of achievement measurement". *Proceedings of the 1969*

- Invitational Conference on testing problems*, pp. 53-61. Princeton, NJ: Educational Testing Service.
- Hartmann, W. (1995). Different Approaches to Comparative Mother-Tongue Education (CME). In W. Bos & R. Lehmann (Eds.), *Reflections on educational achievement*, (pp. 119-136). Münster: Waxman.
- Hertitz, W., Ongstad, S., & van den Ven, P. (Eds.). (2007). *Research on mother tongue education in a comparative perspective. Theoretical and methodological issues*. Amsterdam: Editions Rodopi.
- IEA IPS. *Indagine sulla produzione scritta: Italia. Ricerca educativa*, 1988, (2-3), 1-251.
- Kaplan, R. (1966). Cultural thought patterns in intercultural education. *Language Learning* 6 (1), 1-20.
- Lamb, H. L. (1987). *Writing performance in New Zealand schools. A report of the IEA study of written composition in New Zealand*. Wellington: Department of Education.
- Löftqvist, G. (1988). *Elever skriver uppsats. [Pupils write composition.]* Lund: Studentlitteratur.
- Lewis, E. G., & Massad, C. E. (1975). *The teaching of English as a foreign language in ten countries*. Stockholm: Almqvist & Wiksell.
- Lindell, E. (1980). *Six reports on free writing: A summary of the Fris project*. Malmö School of Education. Lund Universitet: Department of Educational and Psychological Research, Report No. 61.
- Lloyd-Jones, X. (1977). Primary trait scoring. In C. Cooper, & L. Odell (Eds.), *Evaluating writing: Describing, Measuring, judging*, (pp. 33-66). Champaign, IL: National Council for Teachers of English.
- Odell, L. (1982). Defining and assessing competence in writing. In C. R. Cooper (Ed.), *The nature and measurement of competency in English*, (pp. 95-138). Urbana, IL: National Council of Teachers of English.
- Petrovsky, A. R., & Bartholomae, D. (Eds.). (1986). *The teaching of writing. Eighty-fifth yearbook of the National Society for the Study of Education*, Part II. Chicago: University of Chicago Press.

- Purves, A. C. (1973). *Literature education in ten countries*. Stockholm: Almqvist & Wicksell.
- Purves, A. C. (1979). *Achievement in reading and literature in the secondary schools. New Zealand in an international perspective*. Wellington: New Zealand Council for Educational Research.
- Purves, A. C. (Ed.). (1979). *Papers presented at the Pitsburg symposium on cognition and written language*. March 15. ERIC ED 178918.
- Purves, A. C. & Takala, S. (Eds.). (1982). An international perspective on the evaluation of written composition. *Evaluation in Education: An International Review Series* 5 (3).
- Purves, A. C., Söter, A., Takala, S. & Vähäpassi, A. (1984). Towards a domain-referenced system for classifying composition assignments. *Research in the teaching of English* 18, 385-416.
- Purves, A. C. (1987). The evolution of the IEA: A memoir. *Comparative Education Review* 31 (1), 10-28.
- Purves, A. C. (Ed.). (1988). *Writing across cultures. Issues in contrastive rhetoric*. Newbury Park: Sage.
- Purves, A. C. (1992). *The IEA Study of Written Composition II: Education and Performance in Fourteen Countries*. Oxford: Pergamon Press.
- Purves, A. C. (1992). Reflections on research and assessment in written composition. *Research in the Teaching of English* 26, 108-122.
- Sellin, N. (1995). Partial least squares modeling in research on educational achievement. In W. Bos & R. Lehmann (Eds.), *Reflections on educational achievement*, (pp. 156-267). Münster: Waxman.
- Smith, L. S. (1978). *Investigation of writing assessment strategies*. UCLA: Center for the Study of Evaluation.
- Squire, J. R., & Applebee, R.J. (1969). *Teaching English in the United Kingdom. A comparative study*. Champaign, Ill: National Council of Teachers of English.
- Takala, S. (1983). *Student views of what counts as a good composition: some national findings*. Paper presented at the 7th Annual TESOL Conference, Toronto.

- Takala, S., & Vähäpassi, A. (1987). Written communication as an object of comparative research. *Comparative Education Review* 31, 88-105.
- Thorndike, R. L. (1973). *Reading comprehension education in fifteen countries*. Stockholm: Almqvist & Wicksell.
- Weigle, S. C. (2002). *Assessing writing*. Cambridge: Cambridge University Press.
- Wesdorp, H. (1978). *Evaluatietechnieken voor het moedertaalonderwijs. Een inventarisatie van beoordelingsmethoden voor de steekproef, het begrippeel lezen, de spreek-, luister- en discussie-vardigheid* [Evaluation techniques in mother tongue instruction]. Amsterdam: Uva/RTP