J. Charles Alderson

Diagnosing Foreign Language Proficiency

The Interface Between Learning and Assessment

Diagnosing Foreign Language Proficiency

The Interface between Learning and Assessment

J. Charles Alderson

continuum

Acknowledgements

Danish

John E. Andersen (ADT Leader), Anne Holmen, Esthi Kunz, Marie Linnet (Language Representative, Phase 2), Boris Sondersted, Annette Hagel Sorensen.

Dutch

Ron Oostdam (ADT Leader), Lut Baten, H. van Berlo, B. Bossers, R. Fukkink, M. Jetten, Henk Kuijper (Language Representative, Phase 2), P. Lukassen, José Noijons, E. van Schooten, H. Stortelder.

Mats Oscarson (Language Representative, Phase 2), Rauni Paakkunainen, Leena Skur, Sylvi Vigmo.

The partner institutions represented in the ADTs for the 14 DIALANG languages were the following:

Language	*Institution*
Danish:	Københavns Universitet
Dutch:	Universiteit van Amsterdam
English:	Victoria University of Manchester
Finnish:	Jyväskylän yliopisto
French:	Université Charles de Gaulle – Lille 3
German:	Universität Hohenheim
Greek:	Aristoteleio Panepisimio Thessalonikis
Icelandic:	Háskóli Ísland
Irish:	Ollscoil Luimnigh
Italian:	Centro Europeo dell'Educazione, CEDE
Norwegian:	Universitetet i Bergen
Portuguese:	Universidade de Coimbra
Spanish:	Universitat de Barcelona
Swedish:	Göteborgs Universitet

When in October 1998 the management structure of the project was changed, development work in this area was coordinated by the Test Development Working Group, consisting of José Noijons (CITO), team leader, Ari Huhta and Sari Luoma (both from Jyväskylä). In addition, Gabriella Pavan de Gregorio (CEDE) and Johann Fischer (University of Hohenheim) represented those partner institutions which were responsible for item production.

Paul Meara, University of Swansea, and his team, produced the Vocabulary Size Placement Tests in the 14 languages.

The Project Management Team for Phase 1 in Jyväskylä included Kari Sajavaara (Coordinator), Anu Halvari, Jarkko Hänninen, Eija Hietala, Tuija Hirvelä, Ari Huhta, Tarja Jukkala, Felianka Kaftandjieva, Liisa Kelloniemi, Marleena Kojo, Sari Komulainen, Sari Luoma, Maijaliisa Majamäki, Helen Niskanen, Soile Oikkonen, Anne Räsänen, Leena Skur, Peppi Taalas, Sauli Takala, Timo Törmäkangas, Perttu Venermo, Alan Walton.

The Project Management Team for Phases 2 and 3 in Berlin included Wolfgang Mackiewicz (Coordinator of Phase 2), Sabine Kinzius, Gillian McLaughlin, Ernst Miarka, Astrid Peter, Konrad Stransky, Alan Walton.

Dialang website development and maintenance was undertaken in Finland by the Mediakettu company: Jari Peurajärvi, Jouni Takkinen; and in Berlin by Ernst Miarka.

The following institutions made major contributions to the development of the DIALANG assessment system:

- Johannes Kepler Universität Linz – Austria
- Universität Wien – Austria
- Institut für Theoretische und Angewandte Translationswissenschaft, University of Graz – Austria
- Université Libre de Bruxelles – Belgium
- Handelshøjskolen i Århus – Denmark
- Københavns Universitet, Institut for Nordisk Filologi – Denmark
- Université Charles de Gaulle, Lille 3 – France
- DIE, Frankfurt – Germany
- Universität Hohenheim, Stuttgart – Germany
- Aristoteleio Panepisimio Thessalonikis – Greece
- Háskoli Ísland, Reykjavik – Iceland
- Ollscoil Luimnigh, Limerick – Ireland
- Instittiúid Teangeolaíochta Éireann – Ireland
- CEDE, Frascati – Italy
- SCO-Kohnstamm Instituut; Universiteit van Amsterdam – Netherlands
- Universitetet i Bergen – Norway
- Universidade de Coimbra – Portugal
- Departamento de Línguas e Culturas, Universidade de Aveiro – Portugal
- Generalitat de Catalunya/Escoles Oficials d'Idiomes – Spain
- Göteborgs Universitet – Sweden
- Thames Valley University, London – UK
- Victoria University of Manchester – UK

The following institutions assisted the development of the DIALANG assessment system by participating in the trialling of the language tests:

- Vrije Universiteit Brussel – Belgium
- Århus Kommunes Sprogcenter – Denmark
- Århus Universitet – Denmark
- Helsingin yliopiston kielikeskus – Finland
- Jyväskylän ammatillinen aikuiskoulutuskeskus – Finland
- Kuopion aikuiskoulutuskeskus – Finland
- Edupoli, Porvoo – Finland
- Oulun yliopisto – Finland
- Pohjois-Savo Polytechnic, Kuopio – Finland
- Tampereen yliopiston kielikeskus – Finland
- Université de Technologie de Compiègne – France
- Rijksuniversiteit Groningen – Netherlands
- ESADE – Spain
- Universidad Autónoma de Madrid – Spain
- Lunds universitet – Sweden
- Språk- och kommunikationscenter, Utvecklingsavdelningen, Utbildningsförvaltningen, Stockholm – Sweden

- Université de Lausanne – Switzerland
- Centre de langues, Université de Lausanne – Switzerland

Contents

Introduction

This book addresses an age-old topic, studies of which are still in their infancy, as the topic is under-researched and not well theorized or understood: that of the diagnosis of language proficiency. The word 'diagnosis' is common in discussions in language education and applied linguistics and the term 'diagnostic test' is not unusual in discussions of language testing. Such mentions, however, lack exemplification (much less explanation) and even definitions of the term fail to go beyond the superficial 'identification of strengths and weaknesses and their remediation'. As we shall see, even a rigorous field like foreign language testing lacks examples of diagnostic tests, and indeed is frequently confused as to whether it is even possible to devise tests which can diagnose aspects of foreign language proficiency. Related areas like second language acquisition research are still grappling with the fundamental questions of how foreign and second language proficiency develops, and therefore what underlies such development.

Yet at the heart of teaching and assessing foreign language proficiency lies the need to help learners make progress. If researchers, theoreticians and testers do not know how language proficiency develops, they can hardly claim to be able to help learners develop such an ability.

Clearly learners do develop in their proficiency, although few reach similar levels in a foreign language as they have attained in their native language, at least under conditions of instruction. How is this possible? Recent theories have suggested that the way learners develop is at least as much a function of exposure to the foreign language as it is a function of our theoretical or even practical understanding of what develops and how.

So we are faced with a paradox: learners do learn foreign languages, teachers and language educationalists believe they have some understanding of how learners might be encouraged to learn, and the phrase 'diagnosis in order to improve learning' is frequently used. But we lack descriptions in any detail of what changes as learners develop, there is little research underpinning what descriptions there are, and there is a

remarkable lack of valid diagnostic tests or indeed of any tests that claim explicitly to be diagnostic of foreign language proficiency.

It might therefore seem that this book is over-ambitious in its title and its content. The aim, however, is not to explain exactly how and when to diagnose a learner's foreign language proficiency, but rather to put the need for discussions of diagnosis firmly on the agenda of language testers, language educators, applied linguists and second language acquisition researchers. It aims to start a debate about how diagnostic testing might most appropriately be developed. I argue that the field has neglected to construct diagnostic tests, partly because other forms of testing – in particular high-stakes testing – have dominated the field. I am not alone in making this claim, but I also argue that even those who would concentrate their efforts on understanding classroom assessment procedures have failed to address the need for diagnosis of learners' strengths and weaknesses.

This book is not, however, simply a discussion of the need for better diagnosis. Rather, it is based on attempts to develop computer-based diagnostic tests, through a European Union-funded project known as DIALANG. I concentrate on describing and exemplifying the diagnostic system that has been developed in this project, and on exploring the empirical results to date. DIALANG represents a unique opportunity to begin to come to grips with the notion of diagnosis, to explore the limits of what is known to date, and to develop an agenda for further research into the possibilities of the DIALANG system for helping us understand better the nature of foreign language proficiency, and for developing better, more focused and more theoretically informed assessment procedures that might eventually lead to improved diagnosis of the development of foreign language proficiency.

There are those who claim that they already know how foreign language proficiency develops, in particular aspects of the grammatical or morphological systems of some languages, just as there are others who claim that they have already developed scales of language proficiency that show how learners progress. But both sets of claims are inadequate: the former, second language acquisition researchers, because they only focus on narrow aspects of such a complex topic as foreign language proficiency, involving micro-studies which have nothing to say about macro-issues like proficiency development; and the latter, proficiency testers, because they fail to show how real learners do indeed develop, and thus the claims of their scales lack validation.

Nevertheless, we will have made progress in our understanding of diagnosis and diagnostic testing if the two fields of second language acquisition and language testing begin to engage in debate about the need for and the nature of diagnosis, from both micro- and macro-points of view. The history of collaboration between the two fields is not encouraging: although repeated calls have been made for such collaboration (see Bachman and Cohen, 1998 or Banerjee and Franceschina,

2004), there has been little activity to date. But it is my hope that the discussions in this book will spark debate, controversy perhaps, but above all attempts to understand better what needs to be diagnosed and what can be diagnosed, and may lead to attempts to develop better diagnostic instruments than the ones that currently exist.

Thus it is my hope also that this book might assist the age-old but infant field of diagnostic foreign language testing to take a few faltering steps on the long road to maturity. I believe that DIALANG has risen to the challenge, and has shown us how we might move forward, as well as how much remains to be understood. I believe that further research using the DIALANG system will contribute to that understanding. DIALANG is unique in that it attempts the diagnostic assessment of 14 European languages, not merely one dominant language. Although this book focuses almost exclusively on English, that is simply because there is more data available to date on the performance of the English tests in DIALANG. Yet as DIALANG is freely available over the Internet, it is in principle possible to use it as a tool for the understanding of how proficiency in other languages develops, including Icelandic and Irish, Greek and German, Spanish and Swedish, and hence what can be learned about the diagnosis of proficiency in those languages too.

Of course, computer-based diagnosis is limited in what it can achieve: it can tell us little or nothing about how an ability to speak a foreign language develops. But a limited start to our understanding of such a complex topic is better than no start at all, and it is my contention that it is about time we started to understand.

Chapter 1: Diagnosing foreign language proficiency: a teaching/testing interface

Introduction

It is a commonplace to claim the importance of assessment in language teaching and learning. Teachers need to know what learners already know, what they have learned in the course of instruction over a longer or shorter period and where their strengths and weaknesses are, so that they can plan their instruction appropriately, guide learners on where they need to improve and give feedback to learners. Unfortunately, when teachers give students tests, it usually takes a few days, if not longer, for learners to receive their results, and so the feedback lacks immediate relevance. Moreover, tests made by teachers are often of poor quality, and the insight they could offer into achievement, progress, strengths and weaknesses is usually very limited indeed. In the case of national and international examinations and proficiency tests, feedback may be delayed by several months, and the results are irrelevant to learning needs, with little or no information to help learners understand what they need to do in order to improve.

The type of test that comes closest to being central to learning is the diagnostic test.

What is a diagnostic test?

Language testing handbooks frequently distinguish diagnostic tests from placement, progress, achievement and proficiency tests. The ALTE multilingual glossary defines a diagnostic test thus: '*A test which is used for the purpose of discovering a learner's specific strengths or weaknesses. The results may be used in making decisions on future training, learning or teaching*' (ALTE, 1998). However, diagnostic tests are frequently confused with placement tests. The same multilingual glossary defines 'placement test' as follows: '*A test administered in order to place students in a group or class at a level appropriate to their degree of knowledge and ability*' (*op.cit.*). According to this,

both diagnostic and placement tests appear to be designed to identify what a learner knows in order to decide on future teaching or learning. Placement tests, after all, are intended to group learners in homogeneous groups in order to have a suitable basis for further teaching and learning.

The Davies *et al.* dictionary of language testing describes the use of diagnostic tests as follows:

> *information obtained from such (diagnostic) tests is useful at the beginning of a language course, for example, for placement purposes (assigning students to appropriate classes), for selection (deciding which students to admit to a particular course), for planning of courses of instruction or for identifying areas where remedial instruction is necessary.* (Davies *et al.*, 1999)

In other words, diagnostic tests are used for placement purposes – and thus appear to be identical to placement tests. Indeed, Davies *et al.* continue: '*It is common for educational institutions (e.g., universities) to administer diagnostic language tests to incoming students, in order to establish whether or not they need or would benefit from support in the language of instruction used*' (*op.cit.*). In short, there appears to be no distinction between a diagnostic test and a placement test, at least according to these authors.

It is arguable whether it is indeed the case that 'it is common for universities to administer diagnostic language tests'. Much more common is the administration of some form of post-admissions placement test (for an account of such a placement test and the problems of validating it, see Wall *et al.*, 1996). I have never heard any university claim that what it is doing is diagnosing students' strengths and weaknesses with its placement procedures, since typically what is decided is who needs some form of in-sessional or pre-sessional assistance, and who does not. At most, what is decided is whether a student should take a reading class or a grammar class, a writing class or a listening class or, in the case of US institutions, whether a student should take only a fractional load of academic courses rather than a full academic load. This is rarely known as diagnosis.

Bachman (1990) asserts that

> *virtually any language test has some potential for providing diagnostic information ... Information from language tests can be used for diagnosing students' areas of strength and weakness in order to determine appropriate types and levels of teaching and learning activities ... A placement test can be regarded as a broadband diagnostic test in that it distinguishes relatively weak from relatively strong students so that they can be provided learning activities at the appropriate level. Similarly, a readiness test differentiates students who are ready for instruction from those who are not. A detailed analysis of student responses to the questions on placement and readiness tests can also provide more specific information about particular areas of weakness.* (p. 60)

Alderson *et al.* (1995) define diagnostic tests similarly:

Diagnostic tests seek to identify those areas in which a student needs further help. These tests can be fairly general, and show, for example, whether a student needs particular help with one of the four main language skills; or they can be more specific, seeking perhaps to identify weaknesses in a student's use of grammar. These more specific diagnostic tests are not easy to design since it is difficult to diagnose precisely strengths and weaknesses in the complexities of language ability. For this reason there are very few purely diagnostic tests. However, achievement and proficiency tests are themselves frequently used, albeit unsystematically, for diagnostic purposes. (p. 12)

Thus it would appear that even achievement and proficiency tests can perform diagnostic functions. Indeed, Davies *et al.* claim that '*relatively few tests are designed specifically for diagnostic purposes*'. They explain this as follows: '*It is difficult and time-consuming to construct a test which provides detailed diagnostic information*'. Yet it is somewhat odd to say that few diagnostic tests exist because they are time-consuming to construct – proficiency tests are also difficult and time-consuming to construct, but many exist.

It would appear that we have a problem here: diagnosis is useful, most language tests can be used for diagnosis in some sense, it is common for universities to administer diagnostic tests, and yet diagnostic tests are rare!

To summarize: there is considerable confusion in the literature between placement tests and diagnostic tests. Furthermore, it is frequently claimed that achievement and proficiency tests can be used for diagnostic purposes. Yet it is said that diagnostic tests are rare, and are very difficult to construct. Moreover, there are frequent references in the language testing research literature to, and investigations of, proficiency tests, achievement tests, placement tests and even aptitude tests but diagnostic tests are very rarely referred to or investigated.

The reason for this is not immediately apparent. Clearly it would be useful to have tests which enable teachers and learners to diagnose their strengths and weaknesses. However, linguistic diagnosis might not be as well developed as is desirable because of the lack of attention paid to the subject. High-stakes tests that directly affect people's lives, like university entrance tests or proficiency tests for citizenship or employment, have more obvious potentially negative consequences, and so the quality of such instruments is of paramount importance. Inadequate diagnosis in the context of language education is unlikely to be life-threatening, unlike inadequate medical diagnosis. And so much less attention has been devoted to ensuring the validity and reliability of diagnostic tests in the foreign language field.

The aim of this book is to describe and discuss the nature of diagnostic tests, in order to understand better how they might be constructed and validated.