Towards a Theory of Diagnosis in Second and Foreign Language Assessment: Insights from Professional Practice Across Diverse Fields

*J. CHARLES ALDERSON, TINEKE BRUNFAUT and LUKE HARDING

Lancaster University, Lancaster, UK
*E-mail: c.alderson@lancaster.ac.uk

Diagnostic language assessment has received increased research interest in recent years, with particular attention on methods through which diagnostic information can be gleaned from standardized proficiency tests. However, diagnostic procedures in the broader sense have been inadequately theorized to date, with the result that there is still little agreement on precisely what diagnosis in second and foreign language learning actually entails. In order to address this problem, this article investigated how diagnosis is theorized and carried out in a diverse range of professions with a view to finding commonalities that can be applied to the context of language assessment. Ten semi-structured interviews were conducted with professionals from the fields of car mechanics, IT systems support, medicine, psychology and education. Data were then coded, yielding five macro-categories that fit the entire data set: (i) definitions of diagnosis, (ii) means of diagnosis, (iii) key players, (iv) diagnostic procedures, (v) treatment/ follow-up. Based on findings within these categories, a set of five tentative principles of diagnostic language assessment is drawn-up, as well as a list of implications for future research.

INTRODUCTION

Despite a recent wave of interest in diagnosis in language testing and assessment (Alderson 2005, 2007; Alderson and Huhta 2005; Huhta 2008; Lee and Sawaki 2009) there are very few truly diagnostic second and foreign language (SFL) tests. As far back as 1984, Bejar commented: 'Although there is an increasing demand for diagnostic assessment little guidance exists as to how to conduct such assessments' (1984: 185). The situation appears not to have changed much in almost 30 years. There is only a small number of tests [e.g. DIALANG (Alderson 2005; Alderson and Huhta 2005); DELNA (www.delna. auckland.ac.nz/uoa); DELTA (Urmston *et al.* 2013)] which are purposively designed for diagnostic purposes; that is, where the construct, test items and

testing procedures are informed *a priori* by a working theory of SFL diagnosis. Even these, however, may represent an impoverished view of what diagnostic assessment is capable of.

The scarcity of true diagnostic assessment may be a symptom of a lack of a theory of what diagnosis in SFL assessment actually entails (Alderson 2005). Diagnosis, surely, requires more than just instruments for making the diagnosis (as useful as these may be). It presumably also requires a diagnostician who can make a diagnosis, a system for providing feedback and structure for subsequent treatment. The SFL testing field is in need of a more detailed theory, which can account for the multifaceted nature of diagnosis in language assessment. Once a framework for diagnostic assessment has been sketched, research can begin to explore the various facets of this framework with a view to improving diagnostic assessment practices more generally. The aim of this article is to attempt to provide such guidance by exploring the process of diagnosis in other professional and vocational domains, in the hope of contributing to a theory of diagnosis in SFL assessment.

APPROACHES TO DIAGNOSIS

In SFL testing and assessment, diagnostic assessment is usually characterized as focusing on evaluating learners' strengths and weaknesses (see Davies *et al.* 1999; Alderson 2005). In this sense, tests used for diagnostic purposes are theoretically set apart from achievement tests (which measure what has been learned, usually matched to a syllabus), placement tests (which sort candidates into different levels of some course or programme of instruction), and proficiency tests (which measure language ability according to an underlying theory of language). These distinctions are not always clear in practice, however. Much of the recent research on diagnosis has attempted to extract information on strengths and weaknesses from performance on proficiency tests like the TOEFL iBT, and MELAB (see Jang 2009; Lee and Sawaki 2009a, 2009b; Sawaki *et al.* 2009; Li 2011). This is achieved by applying *ex post facto* 'cognitive diagnosis models'—discrete latent variable models—to tests that have not necessarily been designed with diagnosis in mind, but where performance on particular test items might be associated with strengths and weaknesses with particular 'sub-skills' (de la Torre 2009). These statistical models, recently developed by psychometricians (see Leighton and Gierl 2007; Hueber 2010), would certainly appear to have useful applications in the analysis of data derived from diagnostic tests. However, as Alderson (2010) has commented, these models are not applied to diagnostic tests, but represent attempts to reverse-engineer tests intended for other purposes. The models also rely on the psychological reality of sub-skills, and the notion that single items will test discrete sub-skills and not others, both of which have been questioned in the research literature (see Alderson and Lukmani 1989; Song 2008; Grabe 2009).

In addition to these limitations of the current cognitive diagnosis approach, there are also limitations in the way in which diagnosis has been

conceptualized. Specifically, it is not only a focus on identifying strengths and weaknesses which typifies diagnostic assessments. Diagnostic testing procedures—at least in theory—would have other distinguishing features that set them apart from the other types of SFL tests listed above. For example, Alderson and Huhta (2011) outline the following characteristics of a 'truly' diagnostic test:

- more likely to be discrete-point than integrative, or more focused on specific elements than on global abilities;
- less 'authentic' than proficiency or other tests;
- typically low- or no-stakes;
- involves little anxiety or other affective barriers to optimum performance;
- provides immediate results, or as little delayed as possible after test-taking;
- likely to be enhanced by being computer-based;
- enables a detailed analysis and report of responses to items or tasks;
- gives detailed feedback which can be acted upon;
- leads to remediation in further instruction;
- more likely to focus on language than on language skills;
- more likely to focus on 'low-level' language skills than higher-order skills which are more integrated;
- informed by SLA research, or more broadly by applied linguistic theory as well as research;
- based on content which has been or will be covered in instruction, OR based on a specific theory of language development, preferably a detailed rather than a global theory.

Alderson and Huhta (2011: 32) add that a diagnostic test, in fact, would 'focus more on weaknesses than on strengths'. This list, however, is largely speculative, and constitutes only a potential agenda for diagnostic testing, rather than a set of definitive statements about what is necessary or possible. It also does not adequately account for the diagnostician, whose knowledge and experience plays a central role in many broader contexts in which diagnoses are made (see Büscher *et al.* 2011).

DIAGNOSIS BEYOND LANGUAGE TESTING

While language testing is lacking a comprehensive theory of what diagnostic assessment entails, what procedures it follows, what the knowledge is of diagnosticians and how they are trained, and whether 'treatment' is part of the domain of diagnosis, it has been pointed out (Alderson 2005, 2007) that there are many professional and vocational domains where diagnosis is routinely practised, and that language testing, even applied linguistics as a whole, could benefit from studying how diagnoses are conducted in these domains. Medicine, for example, has a long tradition of diagnosis in which the procedures for gathering diagnostic information have changed drastically

through the ages. According to Mortimer (2009), many medical experts believed that the movements of particular planets control the functioning of certain organs: Mercury controls the brain, Jupiter the liver, and so on. More recent literature in the medical field has developed a theory of medical diagnosis as hypothesis refinement. In this view, 'medical diagnosis is viewed as a two-stage process: medical knowledge is first interpreted in a diagnostic sense; next, observed findings are interpreted with respect to this interpreted knowledge and a given hypothesis, yielding a diagnosis' (Lucas 1997: 169).

Theorizing the nature of diagnosis is not uncommon in other fields. One influential such article is by Reiter (1987), who argues that diagnoses need not be unique: there may be several competing explanations for the same faulty system. The normal approach to discriminating among competing diagnoses is to make system measurements, for example, inserting probes into a circuit or performing laboratory tests on a patient. 'Real world diagnostic settings involve observations [. . . to determine] whether something is wrong and hence whether a diagnosis is called for. [. . .] Intuitively, a diagnosis is a *conjecture* that certain of the components are faulty [. . .] and the rest normal. The problem is to specify which components we conjecture to be faulty' (Reiter 1987: 62–63).

However, some of these theoretical writings are forbiddingly dense and frequently require a non-superficial understanding of the specific field involved, be that medicine, systems analysis, or complex computer models. Such theoretical approaches to diagnosis are interesting, but do not throw much light on how diagnosis in SFL might itself be problematized and theorized. For this reason, it seemed to us to be more fruitful to explore with diagnosticians themselves how they see diagnosis in their various fields and specifically how they describe and explain the practice of diagnosis.

THE STUDY: METHOD

The aim of the current study was to explore the range of approaches to diagnosis across various fields in which diagnosis is common. It is not suggested that this dataset represents all professions where diagnosis takes place, nor that all diagnosis in these fields takes place in the ways described by the participants. Rather, the aim was to use these informants in order to map out some parameters of diagnosis across different fields, and in so doing provide a set of potential options for a more comprehensive theory of diagnosis in applied linguistics, and language assessment, in particular. Interviews therefore presented the most effective methodology for this exploratory study as they yield rich data on a phenomenon, and allow for the immediate follow-up of points that are unexpectedly relevant (Rubin and Rubin 2005; Dörnyei 2007).

Participants

Ten informants were identified in fields where diagnosis regularly takes place. These informants were all known to the researchers, and while they may be characterized as a 'convenience sample', they also represent a wide range of professional contexts. The interviewees had considerable expertise and experience in their field, and were based, or had work experience, in countries such as Australia, Finland, Germany, Hungary, Japan, Kenya, Sweden, and the United Kingdom. Table 1 gives an overview of the participants' professions, length of work experience, and their employment context.

Interview procedure

The interviews were semi-structured, with a set of common questions at the beginning and the end of each interview, and a more unstructured approach through the middle stage. All interviews were conducted by a single researcher adhering to the following procedure. First, the interviewer gave a broad explanation of our interest in diagnosis in general and our rationale for approaching representatives of a range of fields. Next, the interviewees were asked to explain how diagnosis in their field is defined and practised. Discussion around these points formed the body of each interview, with questions emerging from issues raised during the interview itself, or being raised to clarify the interviewer's understanding of what had been said. If not already covered, the final part of each interview focused on

Table 1: List of participants

Interviewee's profession	Years of experience	Context
Car mechanic	47	Garage
Computer systems support manager	24	University faculty
Oncologist	27	Hospital; Research; University tuition
General practitioner	30+	General practice
Nurse	35+	Hospital
Neuropsychologist	20+	University faculty
Psychologist/dyslexia expert	20+	University faculty
Special needs teacher for English L1 and L2	20+	University + private tuition
L1 literacy subject specialist, headmistress	25	Primary school
L1 literacy intervention teacher	10	Primary school

the interviewee's training in diagnosis, and the role of subject knowledge and experience in diagnosis. In addition, the interviewees were asked for their views on the relationship between diagnosis-treatment-feedback, and how a diagnosis can be validated. The interviews took place in an office or in the participant's residence and were audio-recorded. Each lasted about one hour.

Analysis

The recordings were transcribed by a professional transcription service, and double-checked for accuracy by the interviewer. The transcripts were then coded by two researchers. Due to the exploratory aim of the study, no *a priori* analytic framework was imposed on the data. Instead, the researchers let themes emerge inductively (Dörnyei 2007) in a bottom-up approach to coding. Broad thematic codes were agreed on through collaboration among the researchers and in consultation with the informants. A number of key themes emerged from this approach, and these categories are presented in detail below.

RESULTS

The coding indicated that the different interviews' content centred on the following topics:

1   Definitions of diagnosis
2   Means of diagnosis

    (2.1) Training
    (2.2) Tools
    (2.3) Knowledge sources

3   Key players
4   Diagnostic procedures
5   Treatment or follow-up

The informants' comments on each of these themes are presented below, thereby identifying main points made, discussing similarities and differences between the reported information and views, and illustrating the analyses with quotes from the different interviews.

1. Definitions of diagnosis

Views on what diagnosis entails were quite diverse across the dataset. For the car mechanic and the computer systems support manager, diagnosis essentially means problem solving. This may be linked to an underlying theory of the system at hand (see [1]), or it may be more haphazard (see [2]).

[1]

Computer systems support manager:

...trying to model the problem, trying to understand all the factors and move yourself forward from that...with a theory of what is going on and why it's occurring.

[2]

Car mechanic:

[it's] trial and error.

This understanding of the term diagnosis may be influenced by the client-service provider context of these professions and the nature of the object of diagnosis (technical artifacts).

Other informants used the term diagnosis in a more narrow sense: problem identification. Treatment or solving the problem is perceived as a separate activity from identifying the cause, although treatment can be the consequence of diagnosis and may be influenced by it (a certain diagnosis may determine that treatment or intervention is needed and what it will entail). In addition, the educators associated the term diagnosis with a particular level of formality that was a step beyond simply 'screening' (see [3]).

[3]

Special needs teacher:

I think screening is like...screening is more informal and you're looking for, sort of, general trends in behaviour and experience, and I think diagnosis is more specific, maybe, and certainly it tends to be more quantitative.

For the neuropsychologist, this formal characteristic of diagnosis tended to be connected with the use of tests or standardized measurement instruments:

[4]

Neuropsychologist:

Like...there is a specific speech sound disorder existing, but we wouldn't say that our dyslexics have specific speech sound disorder unless we would use those tests that those people used who have diagnosed this SSD, speech sound disorder.

The teachers indicated that informal forms of diagnosis are also common practice (e.g. observations as part of everyday teaching), but that they would not label these as diagnosis (see [5]).

[5]

L1 literacy subject specialist:

And we adjust our planning for individual children according to our on-going, if you like, diagnosis. But, in that context, if you actually used the word diagnosis, for me it would be a formal systematic diagnosis.

A common element in all informants' accounts was the primary emphasis on 'problems'—that is, on weaknesses rather than on strengths, when diagnosing (see [6]).

[6]

Interviewer:

I get the impression that in your field you don't talk about strengths.

Neuropsychologist:

That's true. Yes.

This would suggest that Alderson and Huhta (2011) were in alignment with broader fields when they suggested that diagnostic language assessment should focus more on weaknesses than on strengths. It is, however, an approach that is at odds with much current mainstream language assessment where a deficit model is avoided (see discussion in Conclusion).

In summary, the common definition of diagnosis in the data set was that it is a formal activity, sometimes linked to theory (though sometimes simply solution-oriented) which focuses on problem identification, and sometimes treatment, and which by nature focuses more on weaknesses than strengths.

2. Means of diagnosis

2.1 Training

All informants referred to pre-service and in-service training that helped them in diagnosing. The interviewees' comments indicate that pre-service training tends to focus primarily on gaining a thorough knowledge base, and typically involves completing a formal programme (academic or vocational). In-service training consists of a combination of continuing professional development programmes and self-initiated training. In this case, the emphasis lies to a larger extent on learning from experience. A more detailed overview of the different types of training is available in the online supplementary material (Supplementary Table A).

An exception to the typical pre-service training was the computer systems support manager who had a degree in a different field (physics) and little conventional training in computing. He had been interested in computers from his early teens and had acquired most of his knowledge

through experimenting and experience. Another observation is that in the medical sectors the pre-service formal training components contain an important 'on the workfloor' observation/practice component, which the interviewees recognized as contributing to the establishment of expertise in diagnosis (see [7]).

[7]

Nurse:

For medical students, it's really important that they [...] have clinical placement in clinical practice, which involves the patients. [...] students come onto the wards in year two. And the whole purpose of that is that they follow something called the Spiral of Learning. So the thought is that the sooner you can get the students onto the wards, learning at the most basic level from the patients, they can build on that year on year, on year. And take things to a deeper level of learning and understanding. [...]. And the whole point is that you have a hospital full of patients with signs, symptoms, and illnesses, in real life Technicolor, compared to reading the books. So actually, there's no finer and better person to teach you, than a patient who actually has the illness and can use the words to describe how they felt, what it looked like [...]. And that's where the recognition of patterns come in.[...] you need to recognise patterns that are forming. You then put those patterns into the knowledge base that you have, and then you come up with a diagnosis.

Those in language education, however, indicated that they received only limited explicit specific theoretical or practical training in diagnosis during their studies, whereas diagnosis is emphasized in medical programmes (see [8]).

[8]

L1 literacy subject specialist [on her BA in Education]:

We looked at the early reading stage [...] And then we looked at early phonics – the problems and misconceptions that arise there [...] We didn't look at anything diagnostically beyond the basic Barking at Print level – so nothing to do with understanding or comprehension or higher order skills at all.

Although the interviewees valued the pre-service training for fundamental knowledge acquisition, they underlined the critical role of in-service training for their diagnostic work. In particular, exchanging information and experiences amongst colleagues, in a formal or informal manner, appears to make up an important part of the in-service development activities. Several such compulsory, team-driven or self-steered initiatives were reported by those active in car mechanics, medicine, primary school teaching, and computer systems support.

The fact that many interviewees reported initiatives they had personally undertaken seems to indicate that considerable responsibility in diagnostic

expertise development lies with the individual professional, and thus may depend on characteristics such as willingness, interest, or inquisitiveness. Interestingly, the nature of these initiatives is quite diverse. For example, those active in the medical sector considered it crucial to establish, develop and record a pattern recognition system based on case experiences. Those with a psychology background and active in special needs reported to turn to academic research to develop their insights.

2.2 Tools

The interviewees' descriptions of their diagnostic practices show that aid is often sought from a range of tools or that the diagnostic process is guided by particular procedures.

In fields such as car mechanics and computer systems support, databases are often consulted during the diagnostic process. These databases have been developed (most often externally, but sometimes also internally) on the basis of a wide range of experiences. The databases are shared within the field, at a cost or for free, and are accessible within the field (e.g. car mechanics database) or by people in general (e.g. Google). They typically include descriptions of issues ('symptoms'), sometimes—but not always—underlying causes, and details of remedies.

All interviewees reported the use of some form of tests or standardized instruments in the diagnostic process, for example technical tests in car mechanics, physical examinations in medicine, or performance tests in language education or psychology. Specific examples from the language education context include the Schonell and Salford reading tests (UK), or the TORCH reading comprehension tests (Australia). In primary school language education, some of the tests are routinely administered as part of governmental literacy development schemes (e.g. the so-called 'Running Records' reading assessments). In the medical and language education fields, self-made notes recording observations are also consulted for diagnosis.

A number of interviewees reported to be required to adhere to particular procedures. Often, these protocols are institutionalized or have been introduced by government agencies with the aim of standardizing and facilitating the handling of complex situations (e.g. in the medical and language education sectors). They prescribe the steps to be followed during the diagnostic process or ways in which reporting needs to be done. Examples from the UK medical sector are the so-called POTTS charts (Physiological Observations Track and Trigger System) which are used to guide the process of monitoring of patients and the need for action to be taken, and the SBAR procedure (Situation, Background, Assessment, and Recommendation) which is used to report an initial diagnosis to a colleague.

A more detailed overview of the external tools used by the different informants is available in the online supplementary material (Supplementary Table B).

2.3 Knowledge sources

All informants reported that both knowledge and experience play a crucial part in diagnosis. Specific knowledge in a particular area is learned through formal study, pre-service and in-service. The car mechanic, for example, emphasized that although the tools available are great aids, these do not replace the expertise and experience of their user; the diagnostician still needs to decide *when* to use the tools, and *how* to make use of the resources at his/her disposal. All interviewees also stated that previous encounters or cases—'having seen it before'—are instrumental to diagnosis. It was acknowledged that memory is an important mediator in linking past experiences with present observations. Those active in the medical sector (the oncologist, GP, and nurse) used the term 'pattern recognition' when emphasizing experience-driven diagnosis (see [7]).

Furthermore, the interviewees active in the computer systems support, language teaching, and medical sectors described their diagnostic approach as often entailing a holistic evaluation of a constellation of 'symptoms' in which different pieces of information are brought together. The oncologist, for example, reported usage of a heuristic approach, combining past memories based on lengthy expert experience with knowledge that is holistic and integrated (see [9]).

[9]

Oncologist:

But very experienced physicians using mainly the heuristic [approach], which means, after a certain level of knowledge, your memory, your knowledge, and your experience form some sort of complex whole and you can immediately recapitulate, from that complex whole, which way to go or what disease to think.

Even when having a range of measures and tests at one's disposal, interpreting results or performances is not always straightforward, particularly when it concerns human-related diagnosis (language learners, patients). Often, previous experiences come into play when drawing conclusions from external measures. For example, the oncologist mentioned access to algorithms, but also having to make probability judgements that are largely informed by experiences with other patients.

The GP and the oncologist remarked that the balance between knowledge and experience changes with time, whereby diagnosticians typically rely more heavily on knowledge and less on experience early on in their career (see [10]).

[10]

Oncologist:

[V]ery experienced clinicians mainly use the heuristic approach. The trainees and the young doctors use the so-called knowledge base approach. When young doctors start, they always start that sort of knowledge based, ''Okay. This is the laboratory alterations. This is the complaint. So I have read my

*book – they are probably these, and these, and these diseases are possible.''
The heuristic approach [...] that the patient enters the room and you know
already without any, how do you say . . . Not rational thinking but . . .*

Interviewer:

Intuitive.

Oncologist:

*Yeah, it's intuitive. Knowing how you get the feeling. So without really
following your own step of logic, how you get that kind of hypothesis, that
what is the problem with the patients, because [...] everything is a complex
network, your experience, your knowledge, what you have seen before, and
so on, and so on.*

In addition, the GP indicated that the balance depends on the problem, with
some issues being determinable by clear-cut or objective criteria and tools, and
others with vague or rare 'symptoms' requiring more experience to diagnose
(see [11]).

[11]

GP:

*There are rare conditions that you might only come across once in a career
and if you actually make that diagnosis and recognise what it is on the basis
of something you may have remembered from 10-20 years before you sud-
denly feel very proud of yourself. There are times when you just get a feel
based on experience of what's wrong and it may be a common condition
that is presenting in an unusual way.*

What becomes clear from the interviewees' views on means of diagnosis is that
it is to a large extent dependent on an interaction between good training, the
availability of resources, and a diagnostician's experience and expertise.

3. Key players

In addition to the informants themselves, the interviewees also identified a
number of other people as participating in or contributing to the diagnostic
process.

Sometimes, a first impetus is given by the key stakeholders, for example, car
owners, computer users, patients, and teachers. Stakeholder descriptions of
their observation of a problem can contain a first very rough diagnosis,
which may trigger a process of verification by the diagnostician (see [12]).

[12]

Special needs teacher:

*I meet them [the students] for the first time, I try and work out, [...] what is
it that I can help them with, what is it they need to develop? [...] I talk to*

them and find out what they think, what difficulties they've experienced, because [...] these are young adults that I'm working with so some of them are quite self-aware [...] and are able to tell me, you know, ''I've always had this difficulty with this''. Some of them are less self-aware, but they say, ''My tutors keep telling me,'' so it's feedback that they've had through school and from their tutor at university. I sometimes look at their writing if they remember to bring it, I look at their writing, because what I have found as well is that self-reporting is not really reliable.

The different interviewees also reported that, directed by an initial assessment of the problem, they sometimes consult with people specialized in a particular area. For example, the car mechanic occasionally seeks advice from brand specialists and the computer systems manager turns to specialists in networks, operating systems or programming. The nurse and the GP referred to the importance of a team of people in diagnosis, typically with different responsibilities or functions. Sometimes this constitutes joint effort, exchanging and discussing observations and experiences to diagnose (see [13]).

[13]

GP:

We have, about every six weeks to two months, an end-of-life meeting specifically where – because towards the end of life all the doctors can be involved and the nurses [...], we have a team meeting that discusses issues that are relevant to discuss so that everybody is empowered to look after the patient. And one of the things we do talk about is how the diagnosis was made because sadly, with cancer, it is often by some unusual route. [...] 'this is a pattern that happened for this patient', helps everybody else recognise the pattern if it comes again. [...] And sometimes one person can't make a diagnosis, it's not clear, but then somebody else sees them and can feed back and say, ''I've just found this out''.

A detailed list of key players in the diagnostic process, as reported by each individual interviewee, is available in the online supplementary material (Supplementary Table C).

4. Diagnostic procedures

During the interviews, the informants sketched the procedures they often follow when diagnosing. These are described in Table 2. It should be noted that the diagnostic process may not always follow the described pattern and may not be linear in nature, but the informants indicated that these procedures generally characterize their approach and are quite commonly used. The procedures may also be considerably regulated, as witnessed by the use of particular protocols in the medical sector (see, for example, the nurse's summary of procedures in Table 2).

As shown in Table 2, the initial steps tend to involve listening (e.g. to a client, patient, or learner who roughly describes an issue) and observing

Table 2: Diagnostic procedures

Interviewees	Diagnostic procedures

Car mechanic	- Listen to a description of the problem by the client
	- Check whether a light has come on:
	
	  ○ If so, search the computer database for the type of problem; go through computer checklist
	
	    ■ Solve the problem following the database advice
	    ■ Test the car
	    ■ If failing, re-diagnose by thinking through the system of connections
	    ■ If further failure; seek help from an electronics or brand specialist
	
	  ○ If not, it is a mechanical problem;
	
	    ■ Trial and error
	    ■ Solve the problem

Computer systems support manager	- Listen to or read a description of the reported problem by the PC user
	- Holistic approach; activate one's professional knowledge and experience
	- Verify whether it is the problem: Is it the problem? Are there clues that something else is the problem?
	
	  ○ If problem identified: check the solution and solve it
	  ○ If problem not identified:
	
	    ■ Go back to the basic principles (go through a mental list)
	    ■ Solve other problems whilst looking for the problem
	    ■ Eliminate other problems
	    ■ Search databases for known errors
	    ■ Consult colleagues
	    ■ If all else fails, reboot the machine completely

Oncologist	- Spot diagnosis: first sight visual observation of patient - Listen to the patient; ask the patient questions - Holistic pattern recognition - Form a hypothesis of the problem - Test the hypothesis against lab results/images or a medical examination

(Continued)

Table 2: Continued.

Interviewees	Diagnostic procedures

	○ If wrong:
	
	  ■ Re-evaluate by exploring other similar diseases, particularly focusing on specific symptoms (not the generic ones)
	  ■ Test the new hypothesis against lab results/images

General practitioner	Two stages:
	
	a. General diagnosis: medical issue with the person (doctor-centred)
	
	  ○ Spot diagnosis: first sight visual observation of patient
	  ○ Listen to the patient
	  ○ Take and read/check patient notes
	  ○ Medical examination (not always conducted)
	  ○ Check lab results/images, hospital letters
	  ○ Look up diseases and symptoms
	
	b. Detailed and contextualized diagnosis to take action (patient-centred)
	
	  ○ Holistic approach: pull everything together, and also address the patient's fears
	
	If necessary, refer to a specialist

Nurse	Use the POTTS chart: Physiological Observations Track and Trigger System - Basic (visual) observations of the patient - Pattern recognition - Form a holistic judgment; observations are not considered in isolation - Follow diagnostic analysis schemes or a mental checklist; go through these in think-aloud manner with a medical team
Neuropsychologist	- Conduct IQ screening - Administer standardized tests, analyse the data; compare the mean of the test scores (often a composite score) with the control group mean; evaluate the Standard Deviation with the control group data to decide on the disorder and its severity

(Continued)

Table 2: Continued.

Interviewees	Diagnostic procedures
Psychologist/dyslexia expert	- Conduct observations - Conduct IQ screening - Administer standardized tests, analyse the data - Do repeated measurement

Special needs teacher English L1 and L2	- Listen to self-reports by the learners; ask learners questions on how they read; discuss learners' completed checklist with them
	- Study the learners' writing
	- If suspicion of dyslexia, refer for assessment by educational psychologists
	- In the case of L2 speakers:
	
	  ○ Let learners read a text in their L1 aloud
	  ○ Judge the fluency and confidence of reading aloud
	  ○ Ask general comprehension questions
	  ○ Ask the learner for a self-report on the activity and in general
	  ○ Conduct memory tests
	  ○ Make notes on performances
	  ○ Pattern recognition; make use of intuition based on experience

L1 literacy subject specialist	Class teacher: - Conduct observations – formal or informal (day-to-day teaching) - Check formative reading recordings to identify learners who need extra help (with school principal) Intervention teacher: - Conduct observations and assessments - Keep notes on performances - If more problems: call in additional advice or support, or refer to Reading and Language Service advisory teachers - Refer to educational psychologist for further screening/diagnosis
L1 literacy intervention teacher	Class teacher: - Conduct Running Record observations and administer tests - Evaluate which strategies a learner can/cannot use

(Continued)

Table 2: Continued.

Interviewees	Diagnostic procedures
	- Check test results to identify learners who need extra help or those that need to be pushed more (with school principal) Intervention teacher: - Conduct comprehension exercises - Keep notes on performances with the aim of additional diagnosis/profiling

(e.g. 'spot' diagnosis in the medical sector, informal observations by teachers). This may or may not lead to a general assessment of the problem, a process that often includes activation of knowledge and experience. From there, different paths may be followed. For example, the diagnostician may call a helpline for more specific problem identification. This could comprise checking databases, consulting specialists, or administering tests. The final step tends to be holistic in nature, that is 'bringing it all together' to come to a conclusion, which could be a specific diagnosis, inability to diagnose, or referral for further diagnosis. For the car mechanic and the computer systems support manager the process (ideally) concludes with solving the problem.

   Whilst describing diagnostic procedures, the interviewees also made comments on the accuracy of diagnoses. Sometimes a diagnosis is straightforward and precise and one has great certainty. For example, the car mechanic commented that one knows the diagnosis is correct when the treatment works, that is the car is fixed. In other cases, the diagnosis is uncertain and involves more subjective judgement. As the GP and the special needs teacher put it: 'things are not black and white' (GP) and 'it's not an exact science' (teacher). The oncologist associated this with the characteristics of the subject/object of diagnosis; when working with human beings (as in the medical and educational sectors), diagnosis is complicated by the fact that the 'human being has individuality, which means that even the same symptoms are present in different way[s].' Even how much a patient is willing or able to share with the diagnostician affects diagnostic accuracy, the GP noted.

   Although in many cases a range of tests are relied upon, and in fields such as oncology the histology can confirm the certainty of an initial diagnosis, in other areas interpreting test results may involve a considerable extent of subjective judgement. This type of comment was particularly voiced by the psychologists. For example, the neuropsychologist commented that cut-offs for test results ('normal' versus 'deviating') can be quite arbitrary and expertise is crucial in this respect. More fundamentally, he also discussed reliability and

validity issues of some tests and data interpretation, but said that 'because we don't have anything else, we have to trust it'. The special needs teacher shared his concern over the quality of the diagnostic methods. The GP further drew attention to the fact that 'there are diagnoses that you cannot prove [...] because there is no test that will give you a scientific proof'.

The informants with a medical background reported that when lacking more objective measures or when there is less certainty, issues such as expertise in pattern recognition, awareness of the strengths and limitations of tests, and knowledge of the exceptions (e.g. rare diseases) not only influence diagnosis, but also its accuracy. Specialist knowledge and experience were both considered crucial for accurate diagnosis by these informants. The GP added that a diagnostician needs to know his or her own level of competence in order to decide when to seek further advice, and that, apart from developing confidence as a diagnostician over time, it is advantageous to have what he called 'an innate ability to be reasonably confident and reasonably able to cope with uncertainty'. The nurse similarly pointed out that one's personality may play a role in diagnosis, and that a way to minimize inaccuracy or deal with uncertainty is to conduct repeated diagnoses.

This importance of accuracy checks was also brought up by the L1 literacy intervention teacher:

[14]

L1 literacy intervention teacher:

Like a lot of diagnoses [...] on children are based on one assessment, whereas you need constantly to be monitoring.

5. Treatment or follow-up

Different conceptualizations of the relationship between diagnosis and treatment emerged from the interviews. In client–service provider contexts, such as those of the car mechanic and the computer systems support manager, identifying and solving a problem appear to be very closely connected. For example, the computer systems support manager explicitly stated: 'I see diagnosis as problem-solving.' Emphasis is put on treatment, on fixing the problem. Whether or not the problem is specified prior to or during the treatment process is less of a priority.

The medical informants, however, viewed diagnosis and treatment as separate issues. For example, the GP stated:

[15]

GP:

The purpose of diagnosis is not always to do something about it. Sometimes it is just to be able to tell somebody what it is that is wrong.

Diagnosis is thus not so much defined as problem-solving; instead, it constitutes problem identification.

In some cases, neither precise diagnosis nor treatment is the aim (see [16]).

[16]

GP:

And there are times when you don't bother to make a diagnosis because you know that somebody is already too ill to benefit with any treatment.

The psychologists and language education professionals had similar views to the medics in characterizing diagnosis as problem identification rather than problem solving; treatment is not an inherent part of diagnosis. However, for the educators there is an expectation that in most cases results of diagnostic assessment will be used in some manner to help plan schemes of work for individual learners. Furthermore, similar to the car mechanic and computer systems support manager, the psychologists and educators claimed that treatment was possible without a specific diagnosis. The aim of treatment, however, may be different; problem management rather than problem solving (see [17]).

[17]

L1 literacy subject specialist:

[Y]ou can have treatment without diagnosis – because if you are doing something that is working, it doesn't actually really matter what the diagnosis is. [...] our aim is to overcome whatever it is or to move towards not overcoming but managing whatever it is [...] We can't cure dyslexia, but [...] what the pupil can do is manage that problem within their schooling and day-to-day context, strategies to help them to manage that.

DISCUSSION

This article began by outlining the lack of theoretical understandings of diagnosis in the field of second and foreign language assessment. In surveying practitioners from outside the field, a number of key findings have emerged which might be usefully applied in beginning to develop such a theory. First, a common definition of diagnosis is that it is a formal activity of evaluation and judgement, which focuses on problem identification, and sometimes problem-solution or management, and which tends to focus more on weaknesses than strengths. The data also suggest that diagnosis in many fields is often supported by specific assessment tools, training and ongoing professional development, and is greatly enhanced by individuals' diagnostic experience and the involvement of other stakeholders (whether colleagues or patients/students). Procedures for diagnosis varied according to profession, but always involved listening or observing as a first stage, followed by an initial assessment of the

problem (hypothesis forming), then the use of tools, databases, intuition, or specialized help, before the decision-making stage, which required the synthesis of various knowledge strands, sometimes leading to recommendations of treatment or problem resolution.

One of the most noteworthy points to emerge from the interviews is the prevalence of an experiential/intuitive approach to diagnosis among those interviewees who deal with highly complex and dynamic systems (e.g. the GP and the oncologist). This can be compared with the step-by-step, tool-reliant approach used in more mechanistic diagnostic approaches (e.g. the computer systems support manager, the car mechanic). We would argue that SFL diagnosis, relating as it does to the complex and dynamic systems involved in language acquisition, would be more analogous to the context of dealing with the workings of the human body or mind. In this case, it is interesting to note that the more mechanistic approaches which have been advocated to date in, for example, Cognitive Diagnostic Assessment may be useful only inasmuch as they provide a particular type of evidence on which to make a diagnostic decision. Importantly, though, a comprehensive theory of diagnosis in SFL also needs to fully account for the *diagnostician*—their expertise and knowledge, their training and access to other resources, and their behaviour in synthesizing various types of evidence at the decision-making stage.

Tentative principles for diagnostic SFL assessment

In attempting to draw out specific applications for our field, the findings of this study can help to inform a set of five principles which may themselves be understood as a tentative blueprint for diagnostic assessment practices:

1 The first principle that follows from the interviews is that it is not the test that diagnoses, it is the user of the test. This will often be a teacher, who will need to make an informed diagnosis through a process of listening/observation (leading to an initial assessment), then utilizing a range of assessment tools as well as their informed judgement and the expertise of others, and finally forming a decision about the nature of a specific problem. This responsibility of the professional clearly transpired from the interviewees' discussion of their own role and training, and their need to interpret information from a range of sources, whilst also relying on others' observations and expertise (see Results sections 2, 3, and 4).

2 The second principle is that instruments themselves should be designed to be user-friendly, targeted, discrete and efficient in order to assist the teacher in making a diagnosis. Diagnostic tests should be suitable for administration in the classroom, designed or assembled (with recourse to existing suites of tools) by a trained classroom teacher (or other experienced language professional), and should generate rich and detailed feedback for the test-taker. Most importantly, useful testing instruments need to be designed with a specific diagnostic purpose in mind. This principle is

derived from the emphasis the interviewees placed on tools with a clear focus and capacity to play a facilitating role, and thus on tools possessing the above characteristics (see Results section 2).

3 The third principle is that the diagnostic assessment process should include diverse stakeholder views, including learners' self-assessments. As the interviewees indicated, a range of people contribute to the process, which is often initiated by key stakeholder's informal observation of a problem (see Results sections 3 and 4).

4 The fourth principle is that diagnostic assessment should ideally be embedded within a system that allows for all four diagnostic stages: (1) listening/observation, (2) initial assessment, (3) use of tools, tests, expert help, and (4) decision-making (see Results section 4). Much current diagnostic testing arguably begins at stage (3), using general diagnostic tests for whole populations rather than more targeted measures that have been selected on the basis of stages (1) and (2). There is a role to play for large-scale diagnostic tests of this kind (particularly in post-entry language assessment for university programmes), and tailored diagnostic assessment as suggested above will be much more difficult to implement in these contexts. A theory of diagnosis should not preclude large-scale assessments, but it should also pose a challenge to these programmes: would the same decisions about strengths and weaknesses have been made on the basis of an individualized assessment in a classroom context?

5 The fifth principle is that diagnostic assessment should relate, if at all possible, to some future treatment. Although in some fields treatment is inherent to diagnosis whereas in others it is a separate phase, the different interviewees emphasized the need for some sort of remedial action (see Results section 5). The exception to this is the medical field, where diagnosis does not necessarily lead to treatment, because there may be no known treatment, or because the diagnosed condition is too far advanced for anything more than palliative care to be of use. In short, diagnostic language tests should lead to intervention, and the envisaged intervention or treatment should be teachable, or the action to be taken by the learner should be capable of leading to an improvement in that learner's performance.

Implications for research

Implications for research emerge naturally in connection with the five principles listed above:

1 If the teacher is considered a diagnostician, then it is imperative that teachers are given sufficient training, and develop a sufficient knowledge base, to be able to make informed diagnostic assessments. Research

therefore needs to gauge the best method by which to prepare teachers for diagnostic work (either pre-service or in-service). This will necessarily involve the development of a second language acquisition knowledge base, as well as familiarity with a range of tools available for making diagnoses of learner development.

2 The second principle relates to the development and dissemination of well-designed, valid assessment tools. This has been the focus of some large-scale projects, for example DIALUKI, which is exploring a range of linguistic, cognitive and motivational variables which predict strengths and weaknesses in second or foreign language reading and writing, in order to develop diagnostic tools (see www.jyu.fi/hum/laitokset/solki/tut kimus/projektit/dialuki/en). The ongoing development of testing instruments which target specific, atomistic aspects of language knowledge and/or performance is vital for developing a professionalized system of diagnosis. This would ideally result in a repository of free diagnostic tools which would be available to classroom teachers in a similar way to the IRIS digital repository which houses a collection of open-access research instruments for second language acquisition research (see www.iris-data base.org).

3 Self-assessment has been seen as a useful parallel feature of diagnostic assessment for many years (see Spolsky 1992). However, the interview data suggest that self-assessment, as well as the reports of other stakeholders, needs to be integrated into diagnostic decisions in a meaningful way. Research needs to investigate the most reliable ways of reconciling different perspectives in the diagnostic process, and explore how subjective assessments might be combined with the results of objective instruments to create a richer level of insight into particular learning difficulties.

4 Diagnosis should be an embedded process, taking place, wherever possible, in the classroom and feeding back into the curriculum. We therefore need more classroom-based assessment research with a specific focus on the processes of diagnosis (see, e.g. Doe 2011; Fox and Hartwick 2011) as well as a clearer understanding of the interface between diagnosis and treatment in order to map out how this might best be achieved in day-to-day classroom contexts. We might also compare the diagnostic information yielded by large-scale diagnostic tests and the individualized diagnostic approaches suggested in this article.

5 The effectiveness of various treatments or interventions based on diagnostic decisions needs to be investigated. This is, of course, the basis of much research in Second Language Acquisition. However, diagnostic assessment research might focus on the effectiveness of recommended interventions for individual cases—with a consequential focus, in this type of research, on the case study method.

CONCLUSION

In conclusion, while there is still work to be done in articulating a comprehensive theory of diagnosis in SFL assessment, this article has drawn on expertise outside the field to map out some of the broad aspects of what such a theory might entail: a clear definition of diagnosis, a clear understanding of the means of diagnosis and the participants involved, a set of procedures for conducting diagnoses and a closer focus on the interface between a decision and the intervention to follow. We have proposed a set of tentative principles, and based on these principles, a list of research priorities for diagnostic assessment.

There are limitations to the usefulness of surveying diagnosis in wider professions. Several of the professionals interviewed work in fields with a clear normative model on which to base diagnostic decisions (e.g. a healthy human body; a fully functioning computer system). It is much more difficult to locate a clear normative model for second/foreign language development, and this presents further challenges to developing a comprehensive theory of SFL diagnosis. In these professions, diagnosis often also relates to individual objects or subjects. This may connect more closely with diagnosis in the SFL classroom than with large-scale diagnostic enterprises such as DIALANG, DELNA, or DELTA. The principles listed above, however, are of a broad nature, and might be usefully applied in contexts where diagnosis is based on strengths and weaknesses around specific syllabus goals, or alternatively where diagnosis is conducted according to a specific theory of second/foreign language development in a particular skill area (essentially the aim of the DIALUKI project). Furthermore, requiring optimal diagnostic instruments or follow-up (principles 2 and 5) seem sensible regardless of the size of the undertaking. However, more insights may be needed to establish a full-proof theoretical basis for large-scale SFL diagnosis. Potentially, practical implementations of the proposed principles can inform reflections on the characteristics of an encompassing theory.

It should also be added that diagnostic assessment itself needs to be situated within the range of other assessment practices that might routinely take place in and outside the classroom, and we would emphasize that diagnostic assessment is just one type of assessment that provides useful information for students and teachers. With its discourse of 'weaknesses', 'treatments' and 'interventions' there is the risk that diagnostic assessment might lead to a pathologizing of language learning difficulties, and this is not what is intended by the arguments put forth in this article. It may be argued that the particular professions surveyed in this article focus on weaknesses to a degree that is neither useful nor desirable in the context of diagnostic language assessment; we have treated this particular consensus view with caution in our interpretation of the results, and notably the focus on weaknesses has not been included in our five principles. Ultimately, diagnosis is useful for identifying areas where learners need additional help, but it should not be done in isolation from other types of formative assessment that provide important feedback

on strengths. However, because diagnostic assessment is connected with the need for remedial assistance, it is bound to be, in many cases, identified with areas of weakness. As educators, we need not shy away from this, but try to achieve a balance in assessment procedures, and aim to use diagnostic approaches as one element in a repertoire of assessment practices.

SUPPLEMENTARY DATA

Supplementary material is available at *Applied Linguistics* online.

ACKNOWLEDGEMENTS

REFERENCES

Alderson, J. C. 2005. *Diagnosing Foreign Language Proficiency: The Interface between Learning and Assessment*. Continuum.

Alderson, J. C. 2007. 'The challenge of diagnostic testing: Do we know what we are measuring?' in J. Fox, M. Wesche, D. Bayliss, L. Cheng, and C. E. Turner (eds): *Language Testing Reconsidered.* University of Ottawa Press, pp. 21–39.

Alderson, J. C. 2010. 'Cognitive diagnosis and Q-matrices in language assessment: A commentary,' *Language Assessment Quarterly* 7/1: 96–103.

Alderson, J. C. and **A. Huhta.** 2005. 'The development of a suite of computer-based diagnostic tests based on the Common European Framework,' *Language Testing* 22/3: 301–20.

Alderson, J. C. and **A. Huhta.** 2011. 'Can research into the diagnostic testing of reading in a second or foreign language contribute to SLA research?' in L. Roberts, G. Pallotti, and C. Bettoni (eds): *EUROSLA Yearbook 11.* John Benjamins, pp. 30–52.

Alderson, J. C. and **Y. Lukmani.** 1989. 'Cognition and reading: Cognitive levels as embodied in test questions,' *Reading in a Foreign Language* 5/2: 253–70.

Bejar, I. 1984. 'Educational diagnostic assessment,' *Journal of Educational Measurement* 21/2: 175–89.

Büscher, M., D. Goodwin, and **J. Mesman.** 2010. 'Ethnographies of diagnostic work: introduction' in M. Büscher, D. Goodwin, and J. Mesman (eds): *Ethnographies of Diagnostic Work.* Palgrave Macmillan, pp. 1–14.

Council of Europe. 2001. *Common European Framework of Reference for Languages: Learning, Teaching and Assessment*. Cambridge University Press.

Davies, A., A. Brown, C. Elder, K. Hill, T. Lumley, and **T. McNamara.** 1999. *Dictionary of Language Testing*. Cambridge University Press.

de la Torre, J. 2009. 'A cognitive diagnosis model for cognitively based multiple-choice options,' *Applied Psychological Measurement* 33/3: 163–83.

Doe, C. 2011. 'The integration of diagnostic assessment into classroom instruction' in D. Tsagari and I. Csépes (eds): *Classroom-Based Language Assessment.* Peter Lang, pp. 63–76.

Dörnyei, Z. 2007. *Research Methods in Applied Linguistics*. Oxford University Press.

Fox, J. and **P. Hartwick.** 2011. 'Taking a diagnostic turn: Reinventing the portfolio in EAP classrooms' in D. Tsagari and I. Csépes (eds): *Classroom-Based Language Assessment.* Peter Lang, pp. 47–61.

Grabe, W. 2009. *Reading in a Second Language: Moving from Theory to Practice*. Cambridge University Press.

Hueber, A. 2010. 'An overview of recent developments in cognitive diagnostic computer adaptive assessments,' *Practical Assessment, Research and Evaluation* 15/3: 1–7,

available at http://pareonline.net/getvn.asp?v=15&n=3.

Huhta, A. 2008. 'Diagnostic and formative assessment' in B. Spolsky and F. Hult (eds): *Handbook of Educational Linguistics*. Blackwell, pp. 469–82.

Jang, E. E. 2009. 'Demystifying a Q-matrix for making diagnostic inferences about SFL reading skills,' *Language Assessment Quarterly* 6/3: 210–38.

Lee, Y-W. and **Y. Sawaki.** 2009a. 'Cognitive diagnosis approaches to language assessment. An overview,' *Language Assessment Quarterly* 6/3: 172–89.

Lee, Y-W. and **Y. Sawaki.** 2009b. 'Application of three cognitive diagnosis models to ESL reading and listening assessments,' *Language Assessment Quarterly* 6/3: 239–63.

Leighton, J. P. and **M. J. Gierl.** 2007. *Cognitive Diagnostic Assessment for Education: Theory and Applications*. Cambridge University Press.

Li, H. 2011. 'A cognitive diagnostic analysis of the MELAB Reading Test,' *Spaan Fellow Working Papers in Second or Foreign Language Assessment* 9: 17–46.

Lucas, P. 1997. *Lecture Notes in Computer Science*. 1211, 169–180. DOI 10.1007/BFb0029449.

Mortimer, I. 2009. *The Time Traveller's Guide to Medieval England*. Vintage.

Reiter, R. 1987. 'A Theory of Diagnosis from First Principles,' *Artificial Intelligence* 32: 57–95.

Rubin, H. and **I. Rubin.** 2005. *Qualitative Interviewing: The Art of Hearing Data*. Sage.

Sawaki, Y., H-J. Kim, and **G. Gentile.** 2009. 'Q-matrix construction: Defining the link between constructs and test items in large-scale reading and listening comprehension assessments,' *Language Assessment Quarterly* 6/3: 190–209.

Song, M. Y. 2008. 'Do divisible subskills exist in second language (L2) comprehension? A structural equation modeling approach,' *Language Testing* 25/4: 435–64.

Spolsky, B. 1992. 'The gentle art of diagnostic testing revisited' in E. Shohamy and R. Walton (eds): *Language Assessment for Feedback: Testing and other Strategies*. Kendall/Hunt Publishing Company and National Foreign Language Centre, pp. 29–41.

Urmston, A., M. Raquel, and **C. Tsang.** 2013. 'Diagnostic testing of Hong Kong tertiary students' English language proficiency: The development and validation of DELTA,' *Hong Kong Journal of Applied Linguistics* 14/2: 60–82.