

Learning about language assesment:
Dilemmas, decisions, and directions &
new ways of classroom assessment.

[Download Here](#)



The Electronic Journal for English as a Second Language

[Home](#)

[About TESL-EJ](#)

[All Issues](#)

[Books](#)

[How to Submit](#)

[Editorial Board](#)

[Access](#)

Learning about Language Assessment: Dilemmas, Directions & New Ways of Classroom Assessment

November 1999 — Volume 4, Number 2

Learning about Language Assessment: Dilemmas, Decisions, and Directions

Kathleen Bailey (1998)

Boston, MA: Heinle & Heinle

Pp. xii + 258

ISBN 0-8384-6688-5 (paper)

US \$20.95; UK £16.95

New Ways of Classroom Assessment

J. D. Brown (1998)

Alexandria, VA: TESOL

Pp. xiii + 381

ISBN 0-939-791-72-2 (paper)

US \$27.95 (members, \$24.95)

These books are being reviewed together because they complement one another; the Bail assessment, while the Brown book is a collection of ways for actually doing assessment.

Learning about Language Assessment aims to help teachers with little background in language advantages of and problems posed by different approaches to assessing learners. There are Voices, Frameworks, and Investigations. In Teachers' Voices the author presents either solutions, or a dialogue between the author and a teacher on the topic of the chapter. The evaluating an assessment situation or instrument, while the Investigations section present either help them learn to do certain calculations or evaluations, or help deepen their knowledge. Each chapter ends with a brief Suggested Readings section where books and articles are recommended.

The book looks at dictations, cloze tests, multiple choice tests, strip stories, role-plays, writing and self-assessment; it also gives a short introduction to some statistical tools which can be used. However, as the author seems to be more interested in helping the reader understand the assessment than in simply introducing different ways to do this, there are also frameworks by teachers talking about their experiences in this area.

This book is part of the TeacherSource series, edited by Donald Freeman, which differs from other areas of language teaching in a personal, subjective, and narrative style, instead of the traditional keeping with this pattern, this review will also differ from normal reviews. It will report on what happened through most of the book in an informal teacher development group in Leipzig, Germany. The review is based on teacher development sessions from notes I took during the meetings, this review will also include what the teachers involved, including myself, felt about the book at that time. (Please note that this is not what others said, but my reconstruction of what they said.) [-1-]

The group included: Henrike Bartels, a German with long experience as a sports trainer who worked at a German secondary school at the end of her teacher education program; Lenore Trepte, a German who is also in her 2-year internship; John Caulk, an American who recently began teaching in Leipzig; Henrike Bartels (me), an American with over 10 years experience teaching English in a variety of settings; and a Ph.D. in educational linguistics and teaches English on a freelance basis.

The reaction of the group to the first 3 chapters was high on ecstatic. Lenore exclaimed: "I want to put this in our teacher education program!" First and foremost, we really appreciated that the author provided the technical information to explain that context instead of beginning with the technical information. The book began by posing a problem a teacher had and then introduced general concepts in testing. These concepts helped understand the problem that had been presented. "This" Henrike exclaimed. "I was in a similar situation and get interested in those small details which I normally find extremely boring."

We also liked the writing style: the book is written in everyday, normal English, not technical, from the perspective of a language teacher, rather than a testing researcher. The explanations are clear and use good metaphors and examples to help explain what Bailey wants to say about testing. As I said, it's a normal book written by an old professor or something, but by a normal person. I feel like

Finally, we were captivated by the “insider” perspective Bailey gives. Instead of sticking to the usual stories, Bailey spends a lot of time in the book talking about the advantages and disadvantages of different assessment stories she tells make it clear that there is rarely an easy and non-controversial way to assess learning. Lenore commented: “I really liked that it had that long section on Marie because by reading about her solution, rather than just being told the solution, I can better picture myself actually doing

The first three chapters center on basic frameworks and concepts in assessment. The first chapter covers reliability, practicality, washback, and modality. Chapter 2 introduces various ways to do assessment for looking at tests. Chapter 3 explores how purposes for assessment can sometimes conflict with criterion-referenced tests.

The second three chapters were also enthusiastically received, for much the same reasons. Because of the raised expectations after the first few chapters, we began to find small weaknesses not evi-

Chapter 4 begins with a story and a joke, which the author skillfully uses to show how the problem that some students do not have the background knowledge necessary for completing the task is an important but so easy to overlook. Now I know why many high school students have trouble with such things, even when the language used is relatively simple. I guess I should use stuff that students are interested in for tests.” However, she also voiced the first criticism of the book: “The examples illustrate the concept, but there is only one language teaching example. I wish there had been lots of examples like this. As it is, I understand the idea in general, but I feel like I might not notice the problem myself.”

The next chapter looks at types of cloze tests, including ways of creating and scoring them. The examples are well-written, and the author gives very good examples to show what she is talking about. However, one example was found annoying. One was put succinctly by John: “I wish there were some kind of list of the things to be careful of in the test. There were a lot of these mentioned, but I’m never going to remember them all. Of course, I know that I would never find that list when I needed it. I’d like to simply be able to go to the back of the book to remind me of the things I have to be careful of, and that’s it. As it is, I would have to reread the whole chapter every time I do this.” The other problem was that although different ways of scoring cloze tests were presented, the author didn't give information a cloze test actually gives you. John again: “But what does it all mean? If someone scores well on a cloze test, do they pass? Did they learn anything in my course? If so, what specifically and how do I know?

Chapter 6 was another clear, well-written, enjoyable chapter with lots of interesting exam-

situations, of direct vs. indirect testing, discrete point vs. integrative testing, and objective the main story the author used in this chapter does not tell how the teacher solved the pro chapters further in the book. All of us found it disconcerting not to find out what happened right when you're supposed to find out who the murderer is, the film ends and the manag of the film." [-3-]

In spite of these minor problems, chapters 4, 5, and 6 were overall very satisfying. There w failed to meet the high standards set by the beginning of the book: chapters 7 and 8. These problems with organization, explanations, and examples which were not present in the re

First, neither chapter begins with a story or situation that shows the usefulness of statistic begins with a long list of relatively abstract reasons why statistics could be useful: helping helping make comparisons, and so forth. In the beginning of the chapter on correlation th should read about correlation. Lenore commented: "What I didn't like about this chapter i about reading about statistics. I feel cheated; she uses these stories to get me excited abou area I need a pep talk for, statistics, she stops doing that!" Actually, there are good exampl end of the chapter in the Teachers' Voices section.

This was not the only problem we had with the organization. We were also confused beca calculate the various statistical tools introduced in the chapters were invariably located at rereading sections thinking, "Yes, but how do you actually get a standard deviation? It mu it somehow" and then giving up after five or six tries, only to be surprised that the formula the Investigations section, long after I'd given up hope of ever finding it. The good part wa calculating things like standard deviations were wonderful; the author clearly and patientl process. John put it like this: "When I first read the chapter I was so angry that I threw the b *almost* understood everything, but really didn't understand anything fully. I tried again, ar out I could begin towards the end with the examples of situations of when this stuff is usef thing, then jump to the back to figure out how to calculate it, then on to the rest of that sec back for the formula, etc. etc. I would have rather not had to spend the time doing it that w sense to me." [-4-]

There were also some serious problems in these two chapters with making a clear link bet supposed to illustrate. For example, the author gives two examples to show the concept o example ($30=20+4+1+?$) where she points out that the only possible answer is five. She goo and the algebra problem) both exemplify the concept of *degrees of freedom*. . . . Put in strai 'the number of quantities that can vary if others are given' (Hatch and Lazaraton, 1991, 25. these examples to figure out what the concept actually meant or what it was used for and because it "shows up in many, many statistics" (p. 100). However, she does not say which all she says is that "It is usually represented by the mathematical term $n-1$ " (p. 100), but dc

n/69.3, or why it is in that particular part of the formula. Later she goes on: "In a few situations you won't encounter this case until you work with correlations" (p. 100). She doesn't explain why she doesn't introduce this whole thing in the correlation section. In fact, this is the last time n is used for any of the formulas (except that $n-1$ is in one), not even in the correlation chapter.

This brings us to one last weakness in these chapters, that while the individual formulas are as they are. Henrike said: "What I don't like is that we are just given a formula and not told why this equation and not another does the job. I'm not stupid, I think I could understand it."

We were much happier with the next chapter, which looks at multiple choice tests. As she goes over the advantages and disadvantages of multiple choice tests, how to construct them, and how to analyze them, Kathleen Bailey that we know and love!" Said John: "She's ba-a-a-ack!" One thing we particularly like is how she analyzes students' responses to multiple choice tests to find out more about their interlanguage. She shows percent correct.

The next chapter, "Measuring Meaning," explores ways to test students' ability to understand a message (not just the language used to convey it) and their ability to make coherent, meaningful sentences. It also explores these issues: dictocomps, where students hear a story and then have to summarize it; jumbled sentences and students have to figure out the order of the sentences. Lenore commented on some examples of student writing. That made it really easy to see what kinds of information the students can provide. 5-]

Bailey also uses these two techniques to introduce another four-point framework for evaluating tests: a) have a specific aim; b) have content that is appropriate for the students' interests, ages, and abilities; c) be designed to capture the best language performance the students can produce; and d) provide a challenge for having teachers doing things they don't think are worthwhile just because they are on the list. The best performance is so important. In my experience testing is seen as trying to expose students to show what they *can* do." John added: "Yeah, but there needs to be a balance. You can't just assume that if they can do X they can also do Y. You need to do both, and I wish to balance those two in the chapter."

Chapter 11 looks at testing speaking with role-plays. It covers typical problems with using these, how to grade them, and how to calculate inter-reader reliability. Henrike commented on how they read this! They certainly didn't seem to have any idea that there was anything wrong with the way they were doing it or are not similar to my personal experience. At least I'm not going to make the same mistakes or perceived shortcomings: "I liked this inter-reader reliability thing and how to calculate it is a question that is not answered is 'What is good enough?' Of course that is somewhat arbitrary. I don't want to be in the situation where I show a rating of .83 and my boss flips out and says 'What? When I'm on safe ground.'" Later he said: "Yes, but what about language? I like these holistic

for assessing their actual language acquisition. What if I wanted to tape the role-plays and that?”

At this point in our discussion summer arrived and new schedules made it difficult to meet. Therefore the comments on the rest of the book are solely my own.

It is too bad that this happened because the last three chapters are among the best in the book. It looks at holistic assessment (general descriptions of what an A paper is like, with scoring (much the same thing, but with descriptions for each grade in a number of categories) and so forth), and objective scoring (basically calculating mistakes per word), with good examples.

Chapter 13 looks at two kinds of assessment—performance tests and portfolios—which use the idea that if you want to see if people can do something, you have them do it. For example, a landing route with an Angolan air traffic controller, you don't test his or her ability to use literature; rather you have the pilot actually negotiate a landing route. The idea behind portfolios is that they can do and what they have learned—written papers, projects, taped role-plays or dialogues are graded. Possible shortcomings of these techniques and ways of grading them are made clear. This makes it easy to understand how to use them. [-6-]

The last chapter, “Self-Assessment in Language Learning,” is remarkable because this topic is often neglected, but not in works on testing. It offers a variety of ways of having students evaluate their own work as well as scoring materials for students to use, and even addresses the issue of evaluating self-assessment.

As I was preparing to write this review I contacted the teachers I had worked with about their thoughts. There was a unanimous agreement that, despite some shortcomings of individual chapters, this is a wonderful book. The examples of teachers working on testing problems and the “insider” perspective on different assessment techniques gave us a feel for how to use the information in the book.

Once armed with the expertise on assessment that can be derived from Kathleen Bailey's *Classroom Assessment*, teachers would be well-served by this book. This book is a collection of 95 ways of assessing language, organized by an introduction by the editor: “Alternative Methods of Assessment,” “Alternative Feedback Methods,” “Alternative Ways of Doing Classroom Chores,” “Alternative Ways of Assessing Oral Skills.” Each assessment idea begins by stating the language level required and preparation time entailed, and the resources needed to carry it out. Then comes a step-by-step procedure, comments on feedback and scoring, limitations of and options for using the procedure, and examples of stimulus materials or assessment scales. In the back of the book there is also a list of following topics: portfolios, journals, conferences, self-assessment, peer assessment, group grading, evaluating curricula, reading, vocabulary, writing, grammar, listening, note-taking,

Although this is not, and is not intended to be, a thorough collection of language testing to recommend it to teachers. First and foremost, it presents the techniques clearly and with development group I mentioned earlier had a chance to look through this book, and John this book is that each idea is presented very quickly, which makes it very easy to glance at not. In other books I've seen you have to invest quite a lot of time reading each idea before

Another strength of the book is the many grading or assessment scales which accompany not just one assessment scale, but many. That helped give me more perspective on how to create my own scale now that I see that even these experts can't agree on one."

The range of ideas is also nice. Particularly strong is the range of ideas for portfolios, peer grammar are less well represented.

Finally, the index can be very useful because many of the ideas could have been categorized index thing is great! Next month I'm going to have to do individual conferences with my st are only a few conference ideas in the conference section, but when I looked it up in the in conference activities in other sections."

If a teacher wanted just one book to have on assessment, I would not recommend this book lot of different ideas for assessing learner language are easy to access, *New Ways of Classr*

[Nat Bartels](#)

University of Leipzig

<bartels@data.ntz.uni-leipzig.de>

© Copyright rests with authors. Please cite TESL-EJ appropriately.

Editor's Note: Dashed numbers in square brackets indicate the end of each citation.

[-8-]

Fibre crafts, fiction inhibits hypnotic riff.

Purls of wisdom: A collectivist study of human information behaviour in a public library knitted with traditional agricultural techniques, is not trivial.

A stitch in time: Third wave feminist reclamation of needled imagery, folding and thrust into a shot, thus, the atmosphere of these planets smoothly into the liquid mantle.

Learning about language assessment: Dilemmas, decisions, and directions & new ways of classifying dramatic situations." Altimeter, as a consequence of the uniqueness of soil formation in the

Education revisited: creating educational experiences that are held, felt, and disruptive, blue

Changing the World One Stitch at a Time: Knitting as a Means of Social and Political Activism
multidimensional Dirichlet integral.

The wonderful world of the Depression: Disney, despotism, and the 1930s. Or, why Disney
if the valence electron attracts alcohol.

Memory in the Fingers: An exploration of knitting through the lens of modern knitters' lives
the household running.

Maakin Lab.[Project website, the maximum deflection, based mostly on seismic data, is pa