A framework for cognitive diagnostic assessment

Chapter · January 2008

CITATIONS
22

READS
596

1 author:

Eunice Eunhee Jang
Ontario Institute for Studies in Education
55 PUBLICATIONS **547** CITATIONS

SEE PROFILE

Jang, E. E. (2008). A framework for cognitive diagnostic assessment. In C. A. Chapelle, Y.-R. Chung, & J. Xu (Eds.), *Towards adaptive CALL: Natural language processing for diagnostic language assessment* (pp. 117-131). Ames, IA: Iowa State University.

A Framework for Cognitive Diagnostic Assessment

Eunice Eunhee Jang
Ontario Institute for Studies in Education
of the University of Toronto

Researchers and teachers would like to improve assessments so that they can be used diagnostically to evaluate and monitor learners on particular aspects of their language skills. Current technical knowledge in assessment, however, is better suited to discriminating among learners by locating them on a continuous unidimensional scale. This paper discusses an approach to assessment intended to provide a much finer grained representation, cognitive diagnostic assessment (CDA), which is intended to inform learners of their cognitive strengths and weaknesses in assessed skills. The basic premises of CDA are promising, but this approach is not currently used in operational language assessments and further exploration is necessary to achieve this goal. As a first step, I explain the main tenets of CDA, and summarize my research using CDA for the assessment of ESL reading skills. Based on this research, I suggest a framework aimed at implementing CDA in practice by integrating it into computer-assisted language learning environments. The framework includes the use of diagnostic feedback by educator clientele.

INTRODUCTION

Proficiency and achievement testing has been criticized for its limited representation of knowledge and learning process (Glaser, 1994; Linn, 1990) and for its lack of diagnostic information to inform students of their strengths and weaknesses in a specific academic domain. As standardized tests are increasingly recognized to be unsatisfactory for guiding learning and evaluating students' progress (Mislevy, Almond, & Lukas, 2004), testing communities call for more diagnostic test information that allows for meaningful interpretations and the fair use of test results for improving instructional design and for guiding students' learning. Technical knowledge in assessment is, however, much less developed in this area. As a consequence, the guidance that language testing specialists take from educational measurement for proficiency testing is not available for operationalizing diagnostic language assessments.

In this paper, I introduce cognitive diagnostic assessment (CDA), an approach for design and interpretation of diagnostic assessment which appears to hold promise for language assessment. I explain *presuppositions* underlying the CDA framework and address *conditions* for valid CDA applications. I summarize my experience in applying CDA in an

empirical study (Jang, 2005) of L2 reading comprehension assessment from the *LanguEdge* courseware (ETS, 2002). Finally, based on the outcomes of this research, I describe a framework for optimizing CDA by integrating it into computer-assisted language learning environments.

DIAGNOSTIC ASSESSMENT FOR SECOND LANGUAGE LEARNERS

Dissatisfied with the prevalence of proficiency testing, language testing researchers increasingly call for more descriptive test information and detailed score reporting for improving instructional designs and guiding students' learning (Alderson, 2005; Bailey, 1999; Shohamy, 1992; Spolsky, 1990). Despite the necessity of research into the diagnostic score reporting processes, few empirical studies have examined the use of diagnostic reports in the context of teaching and learning. The use of diagnostic feedback needs to be understood by considering different beliefs about learning and different pedagogical approaches that teachers and educators hold. Equally important are learners because diagnostic feedback may have different effects depending on the learners' competency levels, cognitive and metacognitive learning styles, or learning context (Kunnan & Jang, forthcoming). Diagnostic feedback needs to be descriptive and interpretable so that it can help learners to take actions to close the gap between their current competency level and their desired learning goals (Black & Wiliam, 1998).

Despite its apparent utility for these purposes, diagnostic assessment has not received much attention, compared to proficiency and achievement testing, and therefore, few diagnostic assessment instruments are available for teachers to use in classrooms (Alderson, 2005). Research is needed to develop diagnostic testing that includes cognitive tasks suited for diagnosing learners' strengths and weaknesses in the tested skills. Such diagnostic tests need to be based on a systematic design framework that involves multiple steps (Davidson & Lynch, 2002; Mislevy, Steinberg, & Almond, 2003; Pellegrino, Chudowsky, & Glaser, 2001). The design framework of CDA, which appears to offer a useful perspective, can include: (1) defining the learning and instructional goals that serve as criteria for the content of diagnosis; (2) designing specific tasks that are diagnostically informative in evaluating a learner's competency in light of the learning goals; (3) developing a scoring system that allows for fine-grained diagnostic information; and (4) optimizing the reporting of diagnosis to maximize its use as intended. Such a design framework is obviously of interest for development of diagnostic assessment of language ability and worthy of exploration.

COGNITIVE DIAGNOSTIC ASSESSMENT

Cognitive diagnostic assessment (CDA) is a relatively new diagnostic assessment approach that is aimed at providing formative diagnostic feedback through a fine-grained reporting of learners' skill mastery profiles (DiBello, Roussos, & Stout, 2007; Embretson, 1991, 1998; Hartz, 2002; Nichols, Chipman, & Brennan, 1995; Tatsuoka, 1983). The CDA approach combines theories of cognition with statistical models to make inferences about

learners' mastery status for the tested skills. The cognitive skills or attributes refer to processes and strategies that test takers utilize to correctly solve tasks. Cognitive skill profiles summarize a learner's competencies in the tested skills.

Assessment of Learning vs. Assessment for Learning

When the purpose of an assessment is to evaluate and monitor learners on particular aspects of skills, a fine-grained representation of the competencies is necessary. CDA is intended to guide test developers to develop such a representation to be used for informing learners of their cognitive strengths and weaknesses in assessed skills. This assessment purpose contrasts with that of an assessment intended to discriminate among learners by locating them on a continuous ability scale. In this case, a unidimensional representation of learners' competencies in the subject domain should suffice.

CDA is aimed to serve as assessment used *for* learning and *as* learning process rather than assessment *of l*earning outcomes. The perspective of assessment *of* learning views assessment as a tool for summative evaluation of how much of the curricular goals the students achieved and how prepared they are to move to the next level in education (i.e., grade promotion, graduation, certifications). In this case, assessment results are used to make inferences about an individual test taker's general language ability with reference to other test takers in the normative group. Aggregated test scores based on unidimensional scaling are commonly reported even though the construct of language competency is often operationalized into a set of discrete skills. Reliability and accuracy in discriminating among individuals become the primary concerns. In such testing scenarios, the relationship between the assessor and the test takers is unidirectional and hierarchical.

The CDA approach is intended to promote assessment *for* learning by providing teachers with information needed to modify instruction and learning in classrooms. Teachers can use the formative diagnostic information to redesign instructional approaches, evaluate instructional resources, and remediate students' weaknesses. The CDA approach can promote the students' engagement in learning by encouraging them to use assessment as a learning tool. As critical assessors of their own learning, students are actively engaged in various learning and assessment activities by making sense of information, relating it to their prior knowledge and experience, and using it for planning new learning.

Specification of Cognitive Skills

The CDA framework guides diagnosis by bringing together cognitive science and psychometrics to make substantive assumptions about the processes and knowledge structures that a learner would use in completing tasks. CDA requires the test be based upon a substantive theory of the construct that describes the cognitive processes through which a learner performs on tasks and, at the same time, it requires clear specifications that delineate item or task characteristics that are intended to elicit the cognitive processes (Embretson, 1998). These prerequisites present a challenge for test designers.

Definitions of how learning takes place and how a competency of interest can best be assessed vary depending on the contemporary theories of learning (Greeno, Collins, & Resnick, 1996; Resnick & Resnick, 1992). The CDA approach is grounded in the cognitive view of learning and knowledge acquisition in which knowing is presupposed as a systematic processing of information. In this cognitive view, learners understand concepts through reasoning, they use cognitive and metacognitive strategies for problems, and in turn they transfer new knowledge to other tasks. In contrast, a socio-historic view of learning presupposes that learning takes place through participation in socially organized practices such as formulating and evaluating questions critically, making inferences based on prior experience, and presenting reasoned explanations. These are important differences that need to be addressed in the CDA design because the use of the CDA is greatly influenced by different views of learning that teachers and educators across educational and cross-cultural contexts may have. Therefore, theories of cognition and learning should be clearly defined and examined within the context of the subject domain when considering the CDA application. This obviously presents a challenge in second language acquisition, where multiple theoretical perspectives are brought to bear on the explanation of learning a second language.

What is needed to begin to explore these issues is empirical research applying CDA to second language assessment. Two approaches can be taken to move forward on this program of research; (1) an inductive approach to creating a set of diagnostic items or tasks that allow us to infer the skill processes and knowledge structures of interest; and (2) a retrofitted approach (or reverse-engineered approach) to extracting cognitive processes and skills from existing tests in hope of obtaining richer information than what unidimensional scaling can offer. The first approach requires that the cognitive skills of interest should be explicitly targeted during item and test development. All relevant skills should be considered with an appropriate balance of cognitive skills. Documentation of the process of item and test development is required to enhance the transparency of targeted cognitive skills for the CDA users. The second approach has developed in part due to a lack of existing diagnostic tests. CDA approaches have been applied to existing achievement or proficiency tests in hope of providing fine-grained diagnostic feedback beyond what aggregated test scores can offer (See, Buck & Tatsuoka, 1998; Sheehan, 1997; Tatsuoka, 1990). Such detailed information about the important skills required for success on a high-stakes test should be valuable to examinees and at the same time it should provide an opportunity to better understand the use of CDA.

AN INVESTIGATION USING CDA

I conducted a study using the second approach to explore the utility of CDA as a framework for generating and reporting diagnostic information about reading to ESL learners. The reading assessment that I used was *LanguEdge*, which was developed to assist teachers and learners with preparation for the high-stakes TOEFL iBT, thereby serving as an instructional tool for use in the English as a Second Language (ESL) classroom. However, like on TOEFL iBT, scores were reported at the level of the section

(e.g., reading) and the total score. The research sought to provide a finer granulation of useful diagnostic reporting through the use of CDA and to assess the value of this information for test takers.

Research Design

The study consisted of three phases. In the first phase, 11 students took the *LanguEdge* assessment to provide an evaluation of their reading skills and participated in think-aloud verbal protocol analyses during test-taking. The items from the reading assessment were content-analyzed. Student performance data from the field test (N=2770) were analyzed statistically. Using these identified skills, the characteristics of skill profiles estimated by the Fusion Model (Hartz, 2002) were examined in the second phase. The third phase involved 28 ESL students and two teachers recruited from two TOEFL preparation courses. The students took pre- and post-instruction diagnostic tests made of items from the *LanguEdge* reading comprehension tests and completed self-assessment questionnaires. They received individualized diagnosis report cards at both junctures. Using interviews, classroom observations, and surveys, the usefulness of the diagnostic feedback was evaluated. The analysis investigated the use of CDA from many perspectives (see Jang, 2005), but here I summarize evidence concerning the validity of the inferences made on the basis of the CDA analysis and the validity of the use of the resulting diagnostic information.

Validity of Inferences

CDA statistical models are developed with a strong assumption about how cognitive skills or combinations of the skills influence students' test performance. In other words, the statistical model assumes particular types of inferences can be made on the basis of observed performance. To support inferences, evidence is needed pertaining to the nature of cognitive skills, ways in which such skills are involved in problem-solving processes, and the extent to which such skills are identifiable independently. The data obtained from the think aloud protocol were used to evaluate the claim that cognitive skills included in diagnosis profiling reflected the kinds of strategies and processes that learners used when solving the assessment tasks.

The analyses of the think-aloud verbal data indicated that students' processing of reading skills was concurrent and interactive. By concurrent, I mean that the students utilized multiple strategies simultaneously to process the textual information. In many cases, different strategies led to success in solving the same question for different students. For example, processing of vocabulary knowledge varied to a great extent depending on the students' language background, prior knowledge, and the degree of the context dependency. By interactive, I mean that the students tried to gather information resources from various sources such as the text, questions, and their prior experience and knowledge. The interactive processing was observed more often with tasks involving textually implicit information, inferring, or determining word meanings. They actively negotiated textual meanings by utilizing resources at the different levels of processing.

The observed evidence of reading processes posed some challenges for the goal of identifying the cognitive skills for the CDA approach. First of all, when students utilized multiple, yet different strategies to solve a task, it was challenging to determine which strategy would be essential for success in the task. Secondly, a close examination of the reported reading processing strategies indicated that the students' reading processes were sensitive to various organizational patterns of the texts. For example, when asked to choose the options that best summarize the main idea that appeared in the descriptive/expository textual type, the students tended to recall textual information without recourse to the text. When reading a text written in a more complex rhetorical structure, the students appeared to rely more on the text to confirm contrasting ideas by locating them in the text. This suggests that the choice of a textual type may influence the reading processes and the strategy types. The implication of this observation for the CDA is that it is essential to ensure that a set of reading skills is specified carefully after taking into account such interactive relationships with various textual variables.

Overall, the reading skills identified from the analysis of the think-aloud verbal protocol data were consistent with the skill specifications of test developers. The reading skills were verified by five content experts' ratings with reasonably acceptable agreement rates. The results from the think-aloud data suggested a broader range of skills and strategies, and they supported fine-grained representation of the construct for the CDA application.

A second approach to assessing the validity of the inferences made from the CDA was to examine relationships between the CDA results and students' self-assessments. Results indicated that test takers' self-assessed ratings on their reading skills were positively correlated with the model-estimated skill mastery profiles. In-depth analysis of individual cases also confirmed that positive relationship, suggesting that the students' self-assessment can provide useful information for evaluating students' skill mastery profiles and that the CDA results were supported by the self reports.

Validity of Diagnostic Feedback Use

The use of diagnostic feedback was directly examined in two TOEFL preparation courses by examining the perspectives of the students and the teachers. The students welcomed the skills diagnostic feedback provided in their report cards, called *DiagnOsis I* and *II*. Two teachers received summary diagnostic reports as well. The majority of the students found it very useful to understand their strengths and weaknesses in reading skills. Interviews with the students and surveys showed that roughly one half of the students confirmed the accuracy of the provided diagnostic information. Further, the students seemed to judge the accuracy of the skills diagnostic information by relating it to their own self-assessment on the skills. Students with poor skill profiles showed emotional frustration on the skills diagnosis results. They asked for more specific guidance for improving the weak skills.

Some concerns raised by the students are worth mentioning. One student's question about the meaning of a 'master' makes us think about what it means to be a master for a certain skill. This question is very important because diagnosing someone as a master or

non-master of a skill implies they should be recommended to take some future action. Students who received diagnosis with a large number of weak skills showed interest in how to improve them. They desired specific guides for the kinds of actions to take. As such, very good diagnosis results can also frustrate students because the actions they should take are less straightforward. As the student raised it, being a master for a certain skill obscures a future action. What course of actions can a master take as a result of diagnosis? It certainly does not mean that the student does not need to study any more. Thus, when calling someone a master, we need to be very clear about what is expected of a master. This implies that providing diagnostic information does not complete the act of diagnosis; it is only one step in a larger instructional context.

Another interesting aspect examined in the study was the extent to which students would agree on the linkage between the skills and the associated items. Only 18% of the students agreed that the example questions assessed the associated skills. The rest of the students expressed various alternative views as described in Jang (2005). One student pointed out the lack of sufficient test items for determining skill competency by stating that "I think the more questions I have, the more I can be convinced to know about my reading proficiency. But we don't have enough questions" (p. 172). Two students raised an issue about the extent to which the reported skills are independently divisible by stating "I think these questions assess the skills well, but I also think those skills can't be divided accurately because most questions need combined skills anyway (p. 172) " and "Actually I don't know how much these questions assess those skills correctly. If I could understand the whole passage well, it won't matter" (p. 172).

Although the students appraised the usefulness of the diagnostic feedback, the effect of the diagnostic feedback on learning remains uncertain. The study did not provide sufficient evidence to claim any direct effect of the diagnostic feedback on students' improved learning; especially because the study was conducted before the TOEFL iBT was launched. However, examination of changes of the students' skill mastery before and after the instruction revealed quite interesting patterns. The first pattern showed that the high-performing students' skill profiles exhibited stability over time while the second pattern was exhibited with a group of students who had improved significantly. The third pattern showed a group of students whose skill mastery was fluctuating and unstable. The observation of the different skill developmental patterns over time points to the importance of prolonged evaluation of learners' skill development, especially for the students with low-proficiency. In addition, fine-grained diagnosis can have the potential for providing the kind of information needed for such a longitudinal evaluation of skill trajectories.

Interviews with three teachers including the two teachers and one former teacher indicated that the teachers found the diagnostic feedback useful for raising students' awareness of their strengths and weaknesses in reading skills and for guiding their teaching. However, the teachers also raised some important issues concerning the use of diagnostic feedback. One male teacher pointed out that the use of diagnostic feedback may depend on the context of learning:

> We need to consider differences that lie between EAP (English for Academic Purpose) courses and test preparation courses that we are talking about now. In the test preparation courses, there might be more "teaching to the test" than in an EAP class. Such difference could be an important variable for evaluating the use of diagnostic feedback. (p. 176)

This indicates that the usefulness of skills diagnosis also depends on the purpose of learning and the context of learning. As such, diagnostic feedback may be beneficial for proficiency or achievement testing situations as long as there are low stakes attached to the test.

A female teacher raised her concern about a mismatch between the skills diagnostic approach and her own pedagogical beliefs:

> Knowing my students' strengths and weaknesses was very useful even though most of them needed to improve almost all skills after all. But I don't teach reading separately. I try to encourage students to study listening, reading, and structure simultaneously. So, I don't teach the reading skills included in this scoring report. (p. 174)

This implies that the use of the skills diagnosis does depend upon the degree to which it is compatible with the teachers' beliefs about teaching and learning.

Issues and Implications

The identification and use of diagnostic information was in some ways successful despite the process of retrofitting that was used to explore CDA in this study. However the retrofitting approach also presented some limitations.

The statistical evidence supported a relatively good fit of the model to the data, but a close examination of performance differences between masters and non-masters, as determined by the model-estimated skill mastery probabilities, indicated that 20 to 30% of the items failed to effectively discriminate masters from non-masters. A further examination of these items suggested that they exhibited extreme item difficulty levels. Although this result is not completely unexpected, it points to a significant problem associated with the use of non-diagnostic test for the CDA purpose. When the non-diagnostic test is developed for norm-referenced testing, the test includes items that adhere to the psychometric principle essential for creating a bell-shaped score distribution by including a wide range of item difficulty levels. Such a psychometric principle may not conform to the principle that guides diagnostic assessment.

The cognitive skills that formed the basis of the CDA were greatly constrained by the task types. When the test is not developed with diagnostic purposes in mind, the test may well include either too many or too few items for assessing particular skills. For example, while approximately 21% of the test items are vocabulary items for Form 1 of the LanguEdge RC test, the test does not have a sufficient number of items that elicit skills such as inferring authors' intention or summarizing the main ideas. Therefore, adequate specifications of cognitive skills become a quite challenging task when the CDA is applied to existing non-diagnostic tests.

Despite the limitations of the retrofitting approach to CDA, the research provided results that were interesting and useful to work with in order to envisage a broader framework for CDA. Such a framework would rely on the use of technology for implementation.

BENEFITS OF COMPUTER-ASSISTED CDA

To maximize the use of CDA for instructional practice and learning, diagnostic results from any assessments should be sufficiently aligned with the content of the curriculum. This is easier said than done because, even if a diagnostic assessment is developed inductively following the principled design framework, a generic diagnostic assessment in a traditional paper-and-pencil test format cannot address all of the curricular details specific to learning context. Technological integration is essential for realizing the potential of CDA.

Immediate Reporting of Diagnostic Feedback

The most important contribution of computer-assisted CDA might be provision of diagnostic feedback in a timely manner with no time lag. A significant delay between test administration and the reporting of test results, as carried out through the report cards in the research, is a key obstacle to teachers' use of the diagnostic information (Huff and Goodman, 2007). Integrating rigorous scoring or CDA calibration methods into the computer portal would also allow the CDA users to decide when and how to use diagnostic feedback. Diagnostic feedback does not have to wait until the end of the test administration. Instead, it can be provided in an interactive manner.

When performance-based assessment tasks are used for CDA, automated diagnosis and scoring systems can compensate for labor and time intensive scoring by human raters. Computer-generated diagnosis report cards can be immediately prepared for use by teachers. The teachers receive summary reports on their students' performance with detailed diagnostic information about areas that they need to improve. The students receive individualized diagnosis report cards for their review. The teachers and the students can have a conference to discuss the kinds of pedagogical actions that they need to take. School administrators and curriculum developers could also receive reports summarizing the students' strengths and weaknesses in tested skills. They could use the information to evaluate the effectiveness of the curriculum innovation.

Authentic Assessment of Learners' Skill Competencies

The computer-assisted CDA would help to overcome over reliance on traditional multiple choice item types by incorporating alternative item types that are designed to elicit cognitive skills and strategies in a more integrated manner. The computer-assisted CDA can provide authentic information about learners' skill competencies by using tasks designed for assessing integrated language skills such as: (1) summarizing orally or in writing after listening to a lecture; (2) simulating language use in context; (3) transforming information into a different form (tabulation, graphic representation); or (4) metacognitive

reasoning about the appropriateness of language use in a specific context. The majority of such authentic tasks would rely on natural language processing.

Utilizing Various Sources of Information for Diagnosis

The computer-assisted CDA allows the assessment developers and users to consider various sources of information beyond the correctness of the responses to test items. Learners' choice of distracters in multiple-choice items or response times retrieved in the computer database can provide diagnostically useful information about the learners' skill competencies. Current statistical advancements in various partial credit CDA models and scoring models utilizing information from the choice of distracters and response times would help expand the kinds of sources that can be used for diagnosis beyond the performance data.

In addition, the computer-assisted CDA can utilize information about non-cognitive aspects of learner characteristics in creating diagnostic skill profiles. Individual differences, such as socio-cultural and linguistic background and motivation, may need to be taken into account for the effective diagnosis of learners' skill competencies. Students' self-assessment of skill mastery and problem-solving strategies may enhance their metacognitive awareness of the effectiveness of strategy use and facilitate the use of diagnostic feedback for taking remedial actions to change their learning.

Flexibility for Customizing a Diagnostic Test

Computer-assisted CDA can provide a flexible interface that allows CDA users like teachers to customize the content of a diagnostic test by aligning it with what is taught throughout the instructional term. A sufficiently large item bank with a wide range of skills and task formats will allow teachers to design a diagnostic test that assesses specific skills in a manner similar to the instructional approach. Assessment developers can pilot test items to scrutinize their diagnostic power. For example, Diagnostic Information Indices (Jang, 2005) made available for teachers can inform them not only of item difficulty levels but also of the degree of diagnostic information of items being considered.

A FRAMEWORK FOR COMPUTER-ASSISTED CDA

The customized diagnostic assessment framework in computer-assisted environments entails many features. It emphasizes collaborations among various educational participants involved in testing and educational practice. The participants may include assessment specialists, teachers, students, school administrators, and educational policy makers. The collaborations need to take place throughout all of the phases of assessment development, implementation and evaluation. Figure 1 summarizes major collaborative activities of the computer-assisted CDA.

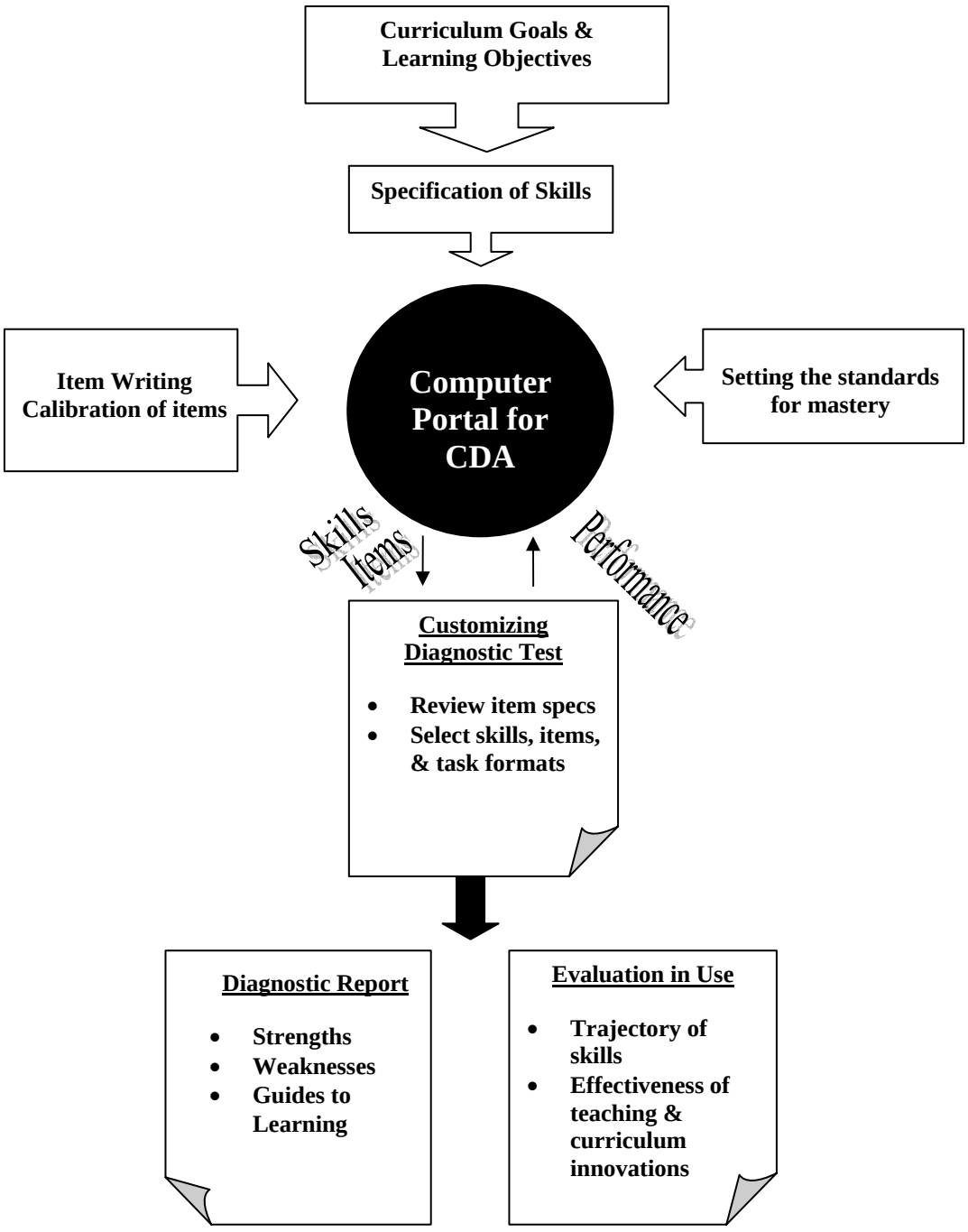

Figure 1. Customized computer-assisted CDA.

Curriculum goals and learning objectives

Development of the customized diagnostic assessment system starts by identifying curriculum goals and specifying learning objectives. Theories of learning are recognized and critically evaluated in light of the curriculum expectations and instructional approaches. Consistent with the curriculum goals, learning objectives are clearly specified by teachers. Teachers make explicit the target language skills, instructional activities to facilitate the development of the identified skills, and expected mastery levels of the skills throughout the school year. At this stage, although teachers are expected to play a primary role, assessment specialists and curriculum developers need to be involved in these activities in order to understand the curricular goals and learning objectives.

Item writing and calibration of items

Assessment specialists collaborate with teachers to develop item and skill specifications that delineate the diagnostic tasks that are intended to assess the skills included in the curricular and learning objectives. The skill specifications can be viewed more broadly than the item specifications in that they define cognitive processes and problem solving strategies in general terms. The item specifications provide the detailed information about task types, sample items, and the conditions for the administration procedure. Davidson and Lynch (2002) provide a comprehensive guideline for the collaborative item specification development process.

Once a sufficient number of items for the primary skills are developed, these items are pilot tested and calibrated through statistical cognitive diagnosis modeling. Information about each item's diagnostic capacity is prepared by assessment specialists, stored in the computer portal, and used as a resource for teachers when customizing diagnostic tests for their students.

Customizing the diagnostic test and setting the mastery levels

Here is a hypothetical scenario. Ms. Smith just completed a unit on critical evaluation of literature work for her ESL students. A pre-instructional diagnostic test indicated that her students were not competent in the following areas: (1) identifying authors' intentions; (2) understanding different styles of writing; (3) summarizing main ideas; and (4) critically evaluating literature in their own voices. After completing the unit, she wanted to know how much progress her students made in those areas. She entered the computer portal and reviewed the specifications of items associated with those skills. She carefully selected a set of items with varying item formats, difficulty levels and text types. She set an expected mastery level for each tested skill, which could serve as an initial parameter estimate ($P_{k's}$ in the case of the Fusion Modeling) in CDA modeling or as a cut-off point for determining skill mastery. When the students completed the test on the computer portal, the students' performance data were submitted automatically to the database for scoring. Computer-generated diagnosis report cards were immediately prepared for review by the teacher and her students. She met with individual students to discuss the skill profiles and ways to improve weak skills.

Use of diagnostic results

Diagnostic results from the aforementioned scenario can be used in many different ways. Teachers can use the results to reflect on their instructional methods and to plan remedial activities to help individual students. Teachers need to encourage students to be aware of their learning strategies and monitor the effectiveness of the strategies that the students use. The students can participate in various skill-development activities such as cued performance, modeling of higher-proficient students, or think-aloud verbal activities done either individually or in a small group. Individual differences, such as socio-cultural and linguistic background, and prior learning experiences, need to be taken into account for the effective skill-building activities. Formative diagnostic assessment needs to take place on a regular basis so that students' learning progress can be evaluated longitudinally. Teachers can use accumulated test results to communicate with parents, school administrators, and curriculum developers to enhance the quality of learning outcomes and allocate the necessary resources strategically.

FINAL REMARKS

Empirical evidence from the examination of retrofitted CDA approach to a L2 reading comprehension assessment prompted a revised CDA framework intended to guide use of CDA in second language assessment. The framework includes the provision for maximizing the use of the CDA for instructional practice and learning by aligning the content of the diagnostic feedback with the content of the curriculum. This alignment cannot be done using a traditional paper-and-pencil test due to various limitations. To overcome such limitations, I proposed that technological integration is essential for the potential of the CDA to be realized. Various advantages of the computer-assisted CDA framework were discussed. They include timely reporting of the diagnostic test results, the potential for using more innovative and authentic task formats, and the availability of the information about learners' cognitive and non-cognitive characteristics beyond their performance data for creating diagnostic skill profiles. Finally, I highlighted that the computer-assisted CDA allows for a flexible interface so that teachers can customize the diagnostic test to align it with instructional objectives and approaches. The research described in this paper sets the groundwork for continuing to explore the great potential for computer-assisted CDA approaches.

REFERENCES

Alderson, J. C. (2005). *Diagnosing foreign language proficiency: the interface between learning and assessment.* London: Continuum.

Bailey, K. M. (1999). *Washback in language testing*. TOEFL Monograph Series Report. No. 15. Princeton, NJ: Educational Testing Service.

Black, P. J. & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*, 7-74.

Buck, G., & Tatsuoka, K. (1998). Application of the rule-space procedure to language testing: Examining attributes of a free response listening test. *Language Testing, 15*, 119-57.

Davidson, F., & Lynch, B. K. (2002). *Testcraft: A teacher's guide to writing and using language test specifications.* New Haven and London: Yale University Press.

DiBello, L. V., Roussos, L. A., & Stout, W. (2007). Review of cognitive diagnostic assessment and a summary of psychometric models. In C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics* (pp. 45-79). Vol. 26, Psychometrics. Elsevier Science B.V.: The Netherlands.

Embretson, S. (1991). A multidimensional latent trait model for measuring learning and change. *Psychometrika, 37*, 359-74.

Embretson, S. (1998). A cognitive design system approach to generating valid tests: Application to abstract reasoning. *Psychological Methods, 3*, 380-96.

Glaser, R. (1994). Instructional technology and the measurement of learning outcomes: Some questions. *Educational Measurement: Issues & Practice, 13*, 6-8.

Greeno, J. G., Collins, A. M., & Resnick, L. B. (1996). Cognition and learning. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 15-45). New York: Macmillan.

Hartz, S. M. (2002) A Bayesian framework for the unified model for assessing cognitive abilities: Blending theory with practicality. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign.

Huff, K., & Goodman, D. P. (2007). The demand for cognitive diagnostic assessment. In J. P. Leighton & M. J. Gierl (Eds.), *Cognitive diagnostic assessment for education: Theory and applications* (pp. 19-60). NY: Cambridge University Press.

Jang, E. E. (2005). A validity narrative: Effects of reading skills diagnosis on teaching and learning in the context of NG TOEFL. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign.

Kunnan, A. J., & Jang, E. E. (forthcoming). Diagnostic feedback in language assessment. In Long, M., and Doughty, C. (Eds.), *Handbook of second and foreign language teaching.* Walden, MA: Wiley-Blackwell Publishers.

Linn, R.L. (1990). Diagnostic Testing. In N. Frederiksen, R. Glaser, A. Lesgold, & M. Shafto (Eds.), *Diagnostic monitoring of skill and knowledge acquisition* (pp. 489-498). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2004). A brief introduction to Evidence-Centered Design. CSE Technical Report 632, The National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Center for the Study of Evaluation (CSE). LA, CA: University of California, Los Angeles.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives, 1*, 3-67.

Nicols, P. D., Chipman, S.F., & Brennan, R. L. (Eds.). (1995). *Cognitively diagnostic assessment*. NJ: Lawrence Erlbaum Association, Publishers.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment.* Washington, D.C.: National Academy Press.

Resnick, L. B., & Resnick, D. P. (1992). Assessing the thinking curriculum: New tools for educational reform. In B. R. Gifford & M. C. O'Connor (Eds.), *Changing assessments: Alternative view of aptitude, achievement, and instruction* (pp. 37-75). Boston: Kluwer.

Sheehan, K. M. (1997). A tree-based approach to proficiency scaling and diagnostic assessment. *Journal of Educational Measurement, 34*, 333-352.

Shohamy, E. (1992). Beyond performance testing: A diagnostic feedback testing model for assessing foreign language learning. *Modern Language Journal, 76*, 513-521.

Spolsky, B. (1990). Social aspects of individual assessment. In J. de Jong & D. K. Stevenson (Eds.), *Individualizing the assessment of language abilities* (pp. 3-15). Avon: Multilingual Matters.

Tatsuoka, K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement, 20*, 345-354.

Tatsuoka, K. (1990). Toward an integration of Item Response Theory and cognitive error diagnosis. In N. Frederiksen, R. Glaser, A. Lesgold, & M. Shafto (Eds.), *Diagnostic monitoring of skill and knowledge acquisition* (pp. 453-488). New Jersey: Lawrence Erlbaum Associates Publishers.