



TABLE OF CONTENTS

LETTER FROM THE EDITORS

ARTICLES

- 5 **Portfolio Assessment: Increasing Reliability and Validity**
By Dale Griffiee
- 18 **Better Tutoring for Non-Native Speakers of English in College Tutoring and Writing Centers**
By Steven Bookman
- 26 **Responding to Self-Selection Bias in Assessments of Academic Support Programs: A Motivational Control Study of Supplemental Instruction**
By Kenneth W. Gattis

JOIN THE CONVERSATION

- 37 **Developmental Theory: The Next Steps**
By Thomas Brothen and Cathrine Wambach

BOOK REVIEW

- 45 **Talking Shop: Authentic Conversation and Teacher Learning**
By James McNamara

PUBLICATION GUIDELINES

NCLCA MEMBERSHIP INFORMATION

NCLCA Officers

PRESIDENT

Mark May

Eastern Illinois University
Learning Assistance Center
600 Lincoln Avenue
Charleston, IL 61920
Phone: (217) 581-6696
cfmsm@eiu.edu

VICE PRESIDENT

Andrea Mericle

Luther College
700 College Drive
Decorah, IA 52101
Phone: (563) 387-1473
mericlan@Luther.edu

CORRESPONDING SECRETARY

Gretchen Marcus

Goucher College
1021 Dulaney Valley Road
Balimore, MD 21204-2794
Phone: (410) 337-6412
gmarcus@goucher.edu

TREASURER

David Reedy

The Learning Center
4240 Campus Drive
Rhodes State College
Lima, OH 45804
(419) 995-8049
Reedy.D@RhodesState.edu

COMMUNICATIONS CHAIR

Jean Marquez

Holy Cross College
P. O. Box 308
Notre Dame, IN 46556-0308
Phone: (574) 239-8373
jmarquez@hcc-nd.edu

PAST PRESIDENT

Joyce Stumpe

Purdue Cooperative Extension Service
155 Indiana Avenue; Suite 301
Valparaiso, IN 46383
Phone: (219) 465-3555
joyce.stumpe@ces.purdue.edu

RECORDING SECRETARY

Susan Busenbark

Rock Valley College
3301 North Mulford Road
Rockford, IL 61114
Phone: (815) 654-4460
S.Busenbark@rvc.cc.il.us

MEMBERSHIP SECRETARY

Mary Knasinski

TARC
University of Wisconsin-Milwaukee
2200 East Kenwood Blvd.
Milwaukee, WI 53201-0413
Phone: (414) 282-8772
mkk2@uwm.edu

PROFESSIONAL DEVELOPMENT

Linda Dixon

Miami University
Bernard B. Rinella, Jr.
Learning Assistance Center
301 South Campus Avenue, Rm. 23
Oxford, OH 45056
Phone: (513) 529-8741
dixonlj@muohio.edu

NCLCA Membership Application

(Journal subscription included)

Name: _____

Institution: _____

Address: _____

Phone: (_____) _____

Fax: (_____) _____

E-mail address: _____

Send application form and a check made out to NCLCA
for \$40.00 to:

Mary Knasinski
TARC
University of Wisconsin-Milwaukee
2200 East Kenwood Blvd.
Milwaukee, WI 53201-0413
Phone: (414) 282-8772
mkk2@uwm.edu

LETTER FROM THE EDITORS

To our readers:

This fall's journal brings together a collection of articles that will help you to strengthen your learning assistance and developmental programs by taking a new look at portfolio assessment, tutoring for non-native speakers of English, and Supplemental Instruction. In addition, we're pleased to publish an article that continues the discussion of a theory of developmental education.

Our first article explores the use of portfolios as assessment tools. Noting that practitioners intuitively find portfolios to be educationally sound, Dale Griffie carefully summarizes the traditional view of reliability and validity and their application to portfolios. He finds that typical portfolios are neither reliable nor valid, and he begins to explore a new concept of reliability and validity while presenting detailed guidelines for implementing an effective and reliable portfolio assessment project.

Next, concerned about the difficulties non-native speakers of English face in higher education, Steven Bookman proposes the need for specially trained "English as a second language (ESL) literacy tutors" to work as peer or writing center tutors alongside other tutors. These literacy tutors would be trained in linguistics and second language acquisition in order to assist ESL students with the academic language demands of college coursework.

Supplemental Instruction (SI) has long been shown to be an effective learning assistance model. However, past studies have not accounted for whether motivation or SI attendance is more influential in affecting grade benefits. In our third article, Kenneth Gattis reports on a study of students participating in Chemistry SI sessions. Motivation is shown to be a determiner in grade performance while SI attendance provides additional grade benefits.

Many in the field of learning assistance and developmental education have been engaged in discussions formulating a theory or theories that underlie our work. In *Join the Conversation* this fall, Thomas Brothen and Cathrine Wambach continue to discuss their proposal for a comprehensive theory of developmental education. They further explore their three unifying concepts, demandingness, responsiveness, and self-regulation, as they affect the development of teaching and learning techniques in college courses and programs.

Finally, we've all remarked on how much we learn from other colleagues, especially when we can talk about a recent teaching experience and compare notes. James McNamara reviews Talking Shop: Authentic Conversation and Teacher Learning, a collection of essays and reports that presents a model for professional development through authentic conversation with other teachers. Various group structures are presented that encourage sustainable conversation groups to aid the professional development of its members.

Martha Casazza
National-Louis University
122 South Michigan Avenue
Chicago, IL 60603
mcasazza@nl.edu

Nancy Bornstein
Alverno College
3400 South 43 Street
Milwaukee, WI 53234
nancy.bornstein@alverno.edu



NCLCA MEMBERSHIP INFORMATION

What is NCLCA?

The mission of the National College Learning Center Association (NCLCA) is to support learning assistance professionals as they develop and maintain learning centers, programs, and services to enhance student learning at the postsecondary level.

What Does NCLCA Do?

- Promotes professional standards in the areas of administration and management, program and curriculum design, evaluation, and research;
- Acts on learning assistance issues at local, regional, and national levels;
- Assists in the creation of new, and enhancement of existing, learning centers and programs;
- Provides opportunities for professional development, networking, and idea exchange through conferences, workshops, institutes, and publications; and
- Offers forums for celebrating and respecting the profession.

How Can I Participate?

The NCLCA Executive Board is anxious to involve as many learning center professionals as possible in achieving its objectives and meeting our mutual needs. Therefore, we invite you to become a member of the National College Learning Center Association. The membership year extends from October 1 through September 30, and annual dues are \$40.00. Membership includes the NCLCA Newsletter and The Learning Assistance Review, discounted registration for the annual NCLCA Conference, workshops, in-service events, and announcements regarding upcoming NCLCA activities. We look forward to having you as an active member of our growing organization.

PUBLICATION GUIDELINES

As an official publication of the National College Learning Center Association, The Learning Assistance Review seeks to expand and disseminate knowledge about learning centers and to foster communication among learning center professionals. Its audience includes learning center administrators, teaching staff and tutors, as well as other faculty and administrators across the curriculum who are interested in improving the learning skills of post-secondary students.

The journal publishes scholarly articles and reviews that address issues of interest to a broad range of learning center professionals. Primary consideration will be given to articles about program design and evaluation, classroom-based research, the application of theory and research to practice, innovative teaching strategies, student assessment, and other topics that bridge gaps within our diverse discipline.

1. Prepare a manuscript that is approximately 12 to 15 pages in length and includes an introduction, bibliography, and subheadings throughout the text.
2. Include an abstract of 100 words or less that clearly describes the focus of your paper and summarizes its contents.
3. Type the text with double spacing and number the pages. Follow APA style (Publication Manual of the American Psychological Association, 5th edition, 2001).
4. Include your name, title, address, institutional affiliation and telephone number along with the title of the article on a separate cover sheet; the manuscript pages should include a running title at the top of each page with no additional identifying information.
5. Submit all tables or charts camera ready on separate pages.
6. Do not send manuscripts that are under consideration or have been published elsewhere.
7. Send four copies of your manuscript to the following address: Nancy Bornstein, Co-Editor, The Learning Assistance Review, Alverno College, 3400 South 43rd Street, Post Office Box 343922, Milwaukee, WI 53234.

You will receive a letter of acknowledgment that your manuscript has been received. The review process will then take approximately three to six weeks at which time you will receive further notification related to your work. If your manuscript is accepted for publication, a computer disk or e-mail transmission will be requested.

ARTICLES

PORTFOLIO ASSESSMENT: INCREASING RELIABILITY AND VALIDITY

By Dale Griffee, South Plains College

Abstract

The use of portfolios as an assessment tool for college composition has been popular since about the mid-1980s, but questions of reliability and validity remain (Pitts, Coles, & Thomas, 2001). The purpose of this article is to summarize the traditional view of reliability and validity and to present ways they can be increased by developmental education practitioners. However, doing so raises questions about the application of traditional understandings of reliability and validity to portfolios. It is argued that a portfolio is a kind of test and, like any test, must pass the muster of validity and reliability. Most articles on portfolio assessment written by practitioners do not report reliability, and the few that do generally conclude portfolio assessment is unreliable, which either suggests portfolio assessment may not be suitable for summative evaluation or that reliability must be rethought. In this article, the role of rater training and the use of multiple raters are examined and found to be factors contributing to portfolio unreliability. Validity is also discussed and lack of clarity on a definition of good writing is found to be central to an understanding of validation. A list of practical actions that can be taken by portfolio raters is given. Finally, the future role of portfolio assessment is discussed. It is concluded that a) the traditional understanding of reliability may not be suitable for portfolio assessment, b) portfolio assessment as currently practiced may not be valid for large-scale, high stakes testing purposes, c) single teacher classroom use of portfolios is or can be both reliable and valid, and d) the use of standardized tests should not be eliminated until we are clearer on portfolio reliability and validity.

Introduction

While portfolios have been applied to academic writing since the 1970s, their use as an assessment tool for college composition has been current only since about the mid-1980s (Hamp-Lyons & Condon, 2000). Despite their increasing popularity as assessment tools, questions about their reliability and validity remain. In broad terms, "a portfolio is a collection of a person's work or

evidence of some artistic, academic, or scholarly activity” (Adams, 1995, p. 568). Portfolios might be collections of an artist’s sketches, an architect’s building plans, or a chef’s created recipes. As applied to educational assessment practice, a portfolio can be viewed as an alternative form of assessment that measures processes inherent in actual classroom learning and teaching (Padilla, Aninao, & Sung, 1996). Since this article is concerned with assessing writing, a portfolio will be defined as a principled collection of writing which may include multiple genres and multiple drafts of each type (Adams, 1995; Falvey & Cheng, 2000). The purpose of this article is to summarize the traditional view of reliability and validity, explore their application to portfolios, and present ways they can be increased by developmental education practitioners.

Key Concepts and Assumptions

Even though portfolios can be classified as alternative forms of assessment, they still share certain characteristics of all assessment instruments. Chief among these is that a portfolio is a test and, as such, must be validated (Hamp-Lyons & Condon, 2000). A test is defined as a procedure for collecting data on a subject’s ability or knowledge (Seliger & Shohamy, 1989). Assessment is usually defined more broadly as gauging educational outcomes (Darling-Hammond, 1994). Evaluation includes both tests and assessment. Brown (1995) says that evaluation includes all the instruments and processes involved in making judgments about an educational program. Formative evaluation is evaluation that takes place during the program and is for the purpose of improving instruction; summative evaluation takes place after the program is finished and is for the purpose of deciding if certain aspects of the program should be kept or changed (Bachman, 1989). Norm-referenced tests (NRTs), or standardized tests, are large-scale tests designed by professionals for the purpose of comparing students (Hambleton & Sireci, 1997). NRTs are not well suited for evaluating specific programs because they are based on the norm group and not any specific curriculum or program. On the other hand, criterion-referenced tests (CRTs) are typically classroom achievement tests designed by teachers to show mastery of material. They do this by allowing students to know what will be tested and encouraging students to study the material. CRTs are not suited for comparing programs but are well suited for evaluating specific programs because they are closely related to the curriculum.

What is Reliability?

It is generally accepted that for any assessment instrument to be valid, it must be reliable. Definitions of reliability typically contain words such as stable, consistent, and dependable. Technically, reliability measures the amount of random error in test scores. Practically, reliability can be seen as the extent to

Library Subscription for *The Learning Assistance Review*

The Learning Assistance Review is a publication of the National College Learning Center Association (NCLCA). It is published twice a year, in the fall and spring.

The journal seeks to expand and disseminate knowledge about learning centers and to foster communication among learning center professionals. Its audience includes learning center administrators, teaching staff, and tutors as well as other faculty and administrators across the curriculum who are interested in improving the learning skills of postsecondary students.

If you would like an annual subscription to The Learning Assistance Review, please mail or fax, on institutional letterhead, your name, address, telephone number, fax number and e-mail address. Please include a check or P.O. number for invoicing. Institutional subscription rates are \$25.00. Send your requests to:

Mary Knasinski
TARC
University of Wisconsin-Milwaukee
2200 East Kenwood Blvd.
Milwaukee, WI 53201-0413
Phone: (414) 282-8772
mkk2@uwm.edu

James McNamara, is an Instructor of Communication in the Instructional Services Department at Alverno College in Milwaukee, Wisconsin.

which we get the same results from the same test administered under similar conditions (see Feldt & Brennan, 1993, for a more complete discussion). Reliability addresses a strong ethical concern in that we want our students to be treated fairly. For example, we would not consider a test fair if the results fluctuated very much because we would not want student results to depend on chance.

There are different types of reliability, but the one that typically concerns us with portfolios is interrater reliability, which is an approximation of agreement of scores that raters (at least two, but sometimes more) each independently give to several portfolios. Scores from each rater are entered into a statistical program and a correlation is requested. The computer program provides an answer in terms of a coefficient which represents a percent of agreement. A correlation coefficient is expressed as a point followed by two numbers, for instance .70 or .86 with 1.00 being perfect agreement. There is no rule for how much agreement is necessary (e.g., how high the coefficient must be), but coefficients in the high 70s, 80s, or 90s are considered acceptable by evaluators.

What is Validity?

Validity is often discussed by educational measurement and test specialists in technical terms, but validation is closely related to ethics, responsibility, and accountability. Whereas reliability is reported as a number, validity is reported as an argument. The argument takes the form of giving and discussing evidence that convinces the reader that the test instrument, a portfolio in our case, measures what those using it think it measures. This raises two important questions. First, where does the validity reside? In other words, what is it about portfolio assessment that is either valid or possibly invalid? Second, where is the starting point in the validation process?

Many teachers believe portfolios are educationally sound because they measure more than one attempt at achievement, and as a result, portfolios have a great deal of validity, especially compared to standardized tests. Do portfolios actually have greater validity compared to standardized tests? This question raises the issue of where the validity is located, and it assumes that validity resides in the type of assessment (e.g., in the portfolio itself). It also assumes validity can be measured and compared and that we can say one type of assessment has more validity than another type. In contemporary thinking about validation, none of these assumptions are accepted (Messick, 1993). In test validation we are not examining the validity of the portfolio instrument, the portfolio content, or even the rater scores, but rather the way we use the information gathered through the portfolio procedure (Bachman, 1990). Simply put, validity does not reside in the portfolio itself or even the resulting scores, but in the soundness of the interpretations proposed for scores from a

test (Cronbach, 1990). It follows from this that validity is not universal, and we cannot say a certain test is once and for all valid. Rather, a particular interpretation is valid for a particular test administration, in a particular place, at a particular time, for a particular group of people, and for a particular purpose.

The second important question is where do we start in the validation process? The answer is we begin by defining what we believe we are measuring, and we do that as carefully and thoughtfully as we can. After all, if validation is giving evidence that our instrument is measuring what we say it is, then we have to first define what that something is.

Is Portfolio Assessment Reliable?

Portfolio studies that report reliability are hard to come by, but from those we have portfolio reliability in the traditional sense is in doubt. Pitts, Coles, and Thomas (2001) report an empirical data study in which eight raters looked at 13 portfolios created by a cohort of students studying to be medical trainers in the U.K. The raters were trained, rated the 13 portfolios, and then examined and rated them again a month later. The kappa statistic used to determine agreement resulted in a coefficient of .34 which was judged to be inadequate reliability. The researchers conclude that if the prime purpose of the portfolio “is as a tool for self-directed learning, it may be more a catalyst to learning where the process is more important than the product” (p. 354). In other words, portfolios are unreliable in the tester’s sense of the word, for summative evaluation, but may be a useful tool for formative evaluation.

Moss, Beck, Ebbs, Matson, Muchmore, Steele, Taylor, and Herter (1992) conducted an experiment in which five graduate students rated ten portfolios written by public high school students by using a checklist and writing extended case studies. Then an additional two graduate student readers independently reviewed all ten portfolios and also wrote brief narratives comparing the ratings done by the first five assessors. The results of the comparison by the second pair of researchers on the ratings done by the first five researchers indicated an interrater reliability that, Moss et al. say by traditional standards, was unacceptably low.

Koretz, Stecher, Klein, and McCaffrey (1994) report partial findings of a large-scale, statewide portfolio assessment project in Vermont for writing and mathematics. They found that “rater reliability was very low in both subjects (writing and math) in the first year of statewide implementation and improved appreciably in 1993 in math, but not in writing” (p. 7). In writing, the portfolio consisted of a single best piece and other pieces of various types. Rescoring by a second rater was done on a random group using Spearman rank-order

In “Reading Lives: Learning about Culture and Literacy in Teacher Study Groups,” Florio-Ruane and Raphael (2001) emphasize the isolation that teachers feel from the world outside of academe. The authors note that often the diversity of current student bodies is seldom matched by similar diversity in the teacher corps, and they take on the question, how can professional development inform participants about culture and diversity? They also wonder how teachers can learn to teach differently from their training in often homogeneous settings. The participating teachers are encouraged to respond to students’ autobiographical narratives with personal narratives of their own. Group discussions try to guide the teachers to produce different readings of the original narratives, moving always towards legitimate moments of comparison and contrast. The authors rightly posit that groups often move too quickly to consensus, trying to remain polite and non-confrontational in the face of difference. If a group can move beyond such differences to become ongoing and self-sustaining, the product becomes much more valuable.

Other chapters continue this movement toward sustainable conversation groups that facilitate the professional development of their members. While the emphasis shifts from informal and voluntary meetings of new teachers who are trying to anticipate what topics will be central to their development in the field to the sharing of narratives with the teacher as both the narrator and protagonist, the “hero of their own tales,” the central concern remains one of constructing authentic conversations that enable and empower the participants. For example, this idea is the foundation for “Connected Conversations: Forms and Functions of Teacher Talk,” written by Cavazos and the Members of Women Educators of Science and Technology (WEST) (2001). They begin by noting that there is a constant need for space for the safe sharing of ideas between women teachers. The conversation group supplies a location for women to share their stories, their “teacher lore,” and be taken seriously (p. 139). As the group attempts to employ the qualities of “care, concern, and connection” that will allow authentic conversations to be meaningful and useful for all, it also ironically identifies several dilemmas that occur because of these qualities. The questions they raise are important for any group which adopts the strategies posed in Talking Shop: How do new members join existing groups? How does one balance the time spent on professional development and personal responsibilities? And finally, how can a group formed to be safe and informative continue to challenge members in new and innovative ways? Even if the reader accepts the editor’s concluding remarks about the value of “good conversation” presented in Talking Shop: Authentic Conversation and Teacher Learning, these questions need to be considered before the implementation of any of the models presented.

While the admitted focus of the text concerns K-12 teachers, all levels of educational professional development could use the strategies and models employed within Talking Shop to enhance an individual's ability to teach his or her students. In fact, since the notion of authentic conversation ironically connects to one of the very things that instructors in higher education often require of their students, the use of the dialogic as a method of learning, the models discussed might be particularly appropriate for college and university faculty. Several of the discussions presented in the text rely on computer technology to facilitate contact; if the methods proposed were employed and adapted by colleges or universities, online conversations might play an even larger role, primarily because they could bring together individuals in common and diverse, but geographically distant, institutions.

As stated before, Talking Shop is a collection of essays and reports in chapter format which provides examples of authentic conversations at work. In the chapter titled "Translating Themselves: Becoming a Teacher Through Text and Talk," Cook-Sather (2001) proposes that as teachers search for a new metaphor that answers questions of identity and purpose, an apt alternative to current images might be the metaphor of translation. This metaphor is appropriate not only for what teachers attempt to do for students, but also for what occurs to teachers as they move from novice to expert. The process involved teachers exchanging letters with students; the former also took part in discussions facilitated by Cook-Sather and others. According to the author, to change one's roles in an educational setting requires translation. For example, new teachers are often not far removed from their students in terms of chronology; the former have memories of what it is to be a high school or college student, but they must undergo a translation as they move to the other side of the lectern. The authentic conversations that occur between new teachers and continuing students seem particularly useful in helping these teachers put the theory of what it is to make the translation from student to teacher into practice.

In Zellermyer's (2001) chapter, "Resistance as a Catalyst in Teachers' Professional Development," the conversations of individuals who volunteered for sessions devoted to professional development for teachers of writing as a process are analyzed. Just as students often do when they pursue assistance, the teachers were initially looking for quick-fix answers; Zellermyer likens them to "consumers, wanting to get tips" about how to conduct classes (p. 41). These teachers often felt like victims of their environments; through their conversations with each other and their ability to reflect on their situations, they were able to pose different points of view and move towards real change. Such professional development breeds the desire to continue to work for change, to continue to think about progressive movements that might be applied to and fostered in others.

correlation which was around .40. The authors conclude that the portfolio system "did not differentiate well between best pieces and the rest of the portfolio" (p. 7).

In Principle, Why is Portfolio Reliability Low?

One cause of low portfolio assessment reliability is that traditional test reliability assumes standardization. Standardization in this context would mean the evaluation instrument does not change, thus the writing ability being measured does not change, administration of the portfolio does not change, and the ratings of those abilities do not change. Writing ability, however, as demonstrated in multiple portfolio drafts does change. Hamp-Lyons and Condon (1993) say that working with portfolio assessment taught them that what they originally thought was a system was really a process. If portfolios contain multiple writing genres each with multiple drafts, then variation is the characteristic of portfolios, not consistency. Traditional concepts of reliability assume consistency. In fact, one definition of reliability is consistency over time and asks the question, to what extent would a student taking a test on a certain day receive the same or similar score a few days later assuming no learning had taken place? Would a rater assign the same or similar score on multiple ratings? Using the metaphor of a shooting range, traditional reliability asks the question, assuming a marksperson, a rifle, and a certain target, how often would a marksperson hit the bull's eye or at least the target? But, what if the target is constantly moving, or even worse, what if the target changes and there are multiple targets? It may be that portfolios have not achieved acceptable reliability because traditional forms of reliability are not applicable.

In Practice, Why is Portfolio Reliability Low?

A portfolio assessment project, especially for teachers doing one the first time, is often a confusing situation. Anything which adds to the confusion adds to lowered reliability and thus lowered validity. Because those conducting the portfolio assessment project function in multiple roles including teaching, rating, and researching, for convenience, I will refer to them as teachers/raters/researchers (TRRs). In most published accounts, TRRs do not calculate, report, or explicitly discuss reliability, but if we read carefully, we can locate potential sources of unreliability. Below are eleven reasons portfolio assessment projects tend toward low reliability:

1. The term "portfolio" is not clearly defined which results in its purpose not being clear. If students are not clear on the portfolio's purpose, they cannot assemble their portfolio in a systematic or reliable way (Arter & Spandel, 1992).
2. No literature review is initiated before a portfolio assessment project.

This means, at the least, that TRRs begin without background knowledge which might alert them to some of the likely problems related to reliability.

3. No pilot study is done. A small pilot study allows TRRs to experience problems and make adjustments. Jumping headfirst into a portfolio project often means a sudden increased workload, which leads TRRs to exhaustion, and may mean that the portfolio papers are not rated uniformly.
4. Full-time and adjunct TRRs are used. Many portfolio assessment projects (Christian, 1993) use a combination of full-time and adjunct faculty as TRRs which adds to a lack of shared understanding. Baumflek, Bloom, Dettmer, DiToro, Friedland, Gooden, Gooding, Richardson, Hill, McQuillan, Mlynarczyk, Percaccio, and Scordaras (1997) studied portfolio assessment and the issue of full-time and adjunct faculty; they found many problem areas which could result in lowered reliability such as time commitment from part-timers who are not compensated, the lack of opportunities for part-timers to absorb rating norms, and the vulnerability part-timers feel in joining regular faculty in candid conversations about teaching and assessment.
5. Rater training is either not done or not done in sufficient depth to fully discuss portfolio criteria and grading standards. Harrison (1995) hints at such problems while Christian (1993, p. 290) writes "To establish a scoring standard, the team would meet to discuss sample papers before the midterm dry run." This is not likely to be enough of an in-depth discussion to reach a consensus of what they are seeking to measure.
6. Categories are changed in the middle of the project. This creates confusion which results in lowered reliability. Christian (1993) found that TRRs prepared a checklist with categories of pass/try again and changed it to finished/not finished, but they failed to revise the checklist and printed the checklist with the previous categories. This resulted in a printed checklist distributed to students with one set of categories while TRRs were using another set of categories in their practice.
7. There is student confusion as to the meaning of the categories. For example, Christian (1993) reports major student problems in understanding the criteria for the categories of "finished" or "not finished."



TALKING SHOP: AUTHENTIC CONVERSATION AND TEACHER LEARNING

Reviewed by James McNamara, Alverno College

Clark, C. (Ed.) (2001). Talking shop: Authentic conversation and teacher learning. New York: Teachers College Press.

Of paramount importance at all levels of academe is the training and continued professional development of teachers. While excellence is always the desired outcome, real life constraints of money and time often limit the amount and scope of development opportunities that both novice and experienced teachers receive. New teaching methods which stress progressive and forward-thinking initiatives such as classrooms which employ more cooperative learning situations and more integrated curricula are now the norm, but teachers are often at a loss as to how to implement these innovations. A new text, Talking Shop: Authentic Conversation and Teacher Learning, offers a new, and seemingly obvious, look at how teachers might use the simple acts of conversation and reflection to become better at their craft. This collection of essays and reports, edited by Christopher M. Clark, Director of the School of Education at the University of Delaware and former professor of educational psychology at Michigan State University, discusses several examples of teachers engaging in "authentic conversation," a term coined to represent types of discussions that rely on teachers sharing "personal narratives of teaching experiences" (p. 6).

The book's underlying thesis is that teachers' professional development "must be led by teachers themselves and be intrinsically satisfying, voluntary, and inexpensive" (p. 4). In fact, several chapter authors express this concern with the limited economic resources of education, a rude awakening that most teachers don't realize until they are actually practicing their profession; often such real life issues are ignored in graduate school or professional development seminars, where ideals and theories supercede practical realities. The authors identify several other current problems that teachers encounter, such as personal safety, for example, and set out to show how the methods employed in authentic conversation might alleviate or offset such concerns.

- (1994). Supplemental instruction: Increasing achievement and retention. New directions in teaching and learning. San Francisco: Jossey-Bass.
- Menges, R. J. (1994). Improving your teaching. In W. J. McKeachie, Teaching tips: Strategies, research, and theory for college and university teachers (9th ed.) (pp. 297-312). Lexington, MA: DC Heath.
- Richardson, R., Fisk, E., & Okun, M. (1983). Literacy in the open-access college. San Francisco, CA: Jossey-Bass.
- Rose, M. (1985). The language of exclusion: Writing instruction at the university. College English, 47, 341-359.
- Wambach, C. & Brothen, T. (2002). The GC Base Curriculum: Description, historical antecedents, theoretical structure and evaluation outcomes. In D.B. Lundell & J. L. Higbee (Eds), Histories of Developmental Education (pp. 73-81). Minneapolis, MN: Center for Research on Developmental education and Urban Literacy, General College, University of Minnesota.
- Wambach, C., Brothen, T., & Dikel, T. N. (2000). Toward a developmental theory for developmental educators. Journal of Developmental Education, 24 (1), 2-4, 6, 8, 10, 29.
8. A lack of understanding and consensus on the part of TRRs as to grading standards produce variability which is one definition of unreliability. For example, Christian (1993) reports that TRRs sometimes emphasized different categories of writing. This resulted in some papers being graded differently. Moss et al. (1992) found that some of the raters emphasized the drafts (process) while others emphasized the final writing (product).
 9. Rater overload occurs. Christian (1993) reports four raters read two hundred essays at least three times.
 10. Portfolios are complex including multiple drafts of multiple genres. A complex portfolio is less reliable than a simple portfolio. Christian (1993) reports that her first semester portfolio contained four types of papers, which she reduced to two types in the second semester.
 11. Students are not cognitively or emotionally prepared for the portfolio project. This leaves them upset and concerned about their grades which in turn can affect their writing. In Christian (1993), the project was abruptly announced at the beginning of the semester, and students expressed concern throughout most of the semester.

The Role Validity Plays

Earlier it was stated that validity begins with a definition of what is being measured, in our case, writing ability. For example, Metzger and Bryant (1993) argue that in setting up a portfolio system, it is necessary to consider what they call basic ideological considerations, one of which is the status of truth, reality, and knowledge for the teacher. This stance is premised on the relationship between pedagogy and theory, namely that “the pedagogy one chooses assumes that the truth and knowledge that students present in their writing comes from a certain source” (p. 282).

Moss et al. (1992) argue that the validity of the conclusions is warranted by the process of data analysis and the transparency of the evidentiary trail. In other words, validation is obtained if readers can clearly see what is being analyzed and how it is being analyzed. In their project, the data analysis was seen in the narratives which resulted from the interplay of the writing samples and the categories. One set of raters looked at the portfolio samples and wrote narratives describing them. Another set of raters analyzed the narratives using a three-step process: the categories were developed, then applied to the writing, and finally a narrative was written. According to Moss et al., there are two occasions for discussion among raters. One is when a second rater reviews the evidence and the assessment rating (e.g., grade, or pass/fail) and agrees or disagrees with the first rater. This is what most writers reporting portfolio

assessment describe as a calibration session, a session in which TRRs get together in a room, read, and rate portfolios for the purpose of discussing and agreeing on rating standards. Portfolio rating sessions can be marathon events. Usually, two TRRs rate and sometimes a third reader is required when the initial two readers cannot agree.

But there is a discussion before that, a discussion which is not stressed by Moss et al. (1992), and that is when the categories are developed. Because a discussion of the meaning of categories at this level is seldom reported, in most portfolio rating sessions the categories are implicitly assumed. These categories, or rather the lack of them, are the primary stumbling block, or Achilles' heel, of portfolio assessment. The problem is getting a group of teachers who are not experienced in stating the beliefs out of which they operate to explicitly state those beliefs. An even larger problem is not disagreeing with the categories, but understanding what they mean. For example, in Moss et al. one of the categories is vision which is defined as "the reader's reconstruction of the writer's semantic intent--the message apparent in the writing" (p. 15). That is the sum total of the explanation. Perhaps in developing this category, the researchers engaged in extended discussion of the theory and values behind this category, but if so, they did not report it, which means that the rater is unaware of what lies behind this category.

At this point, the problem of portfolio assessment becomes clear. In order to engage in a portfolio assessment project, TRRs must either adopt the categories of Moss et al. (1992) which are vague, adopt the categories of somebody else, or create their own. Either way, if TRRs are working with a group, they must use categories to begin an evidentiary trail another reader can follow. Without these basic categories, they cannot evaluate a portfolio, a second rater cannot assess their evaluation by following their evidentiary trail, and validation of their interpretation is not possible.

What We Can Do to Increase Portfolio Reliability and Validity

The following suggestions for increasing portfolio reliability and validity provide general guidelines because specific situations vary from a single classroom teacher wanting to engage in portfolio assessment for her own class to a departmental or college-wide portfolio project.

1. Explicitly define what you mean by a portfolio. Include the purpose for which you plan to use it.
2. Begin a literature search. Since there is extensive literature on portfolio assessment, it may be helpful to form a group of interested

Thomas Brothen, Ph.D., is Morse-Alumni Distinguished Teaching Professor of Psychology and Social Sciences in the General College of the University of Minnesota in Minneapolis, Minnesota.

Cathrine Wambach, Ph.D., is Morse-Alumni Distinguished Teaching Associate Professor of Psychology and Social Sciences and Director of the Office of Research and Evaluation in the General College of the University of Minnesota in Minneapolis, Minnesota.

References

- Bonham, B. (1990). Research on Developmental Education: An interview with James A. Kulik. Journal of Developmental Education, 13(3), 16-18.
- Boylan, H. (1999). Exploring alternatives to remediation. Journal of Developmental Education, 22(3), 2-10.
- Brothen, T., & Wambach, C. (1999). An analysis of non-performers in a computer-assisted mastery learning course for developmental students. Research and Teaching in Developmental Education, 16, 41-47.
- Brothen, T., & Wambach, C. (2000). A research based approach to developing a computer-assisted course for developmental students. In J. L. Higbee & P. L. Dwinell (Eds.), The many faces of developmental education (pp. 59-72). Warrensburg, MO: National Association for Developmental Education.
- Grubb, W. N. (1999). Honored but invisible: An inside look at teaching in the community colleges. New York: Routledge.
- Higbee, J. L. (1996). Defining developmental education: A commentary. In J. L. Higbee, & P. L. Dwinell (Eds.), Defining developmental education: Theory, research, and pedagogy. NADE Monograph (pp. 63-66). National Association for Developmental Education. Eric No: ED394415.
- Johnson, D., Maruyama, G., Johnson, R., Nelson, D., & Skon, L. (1981). Effects of cooperative, competitive, and individualistic goal structures on achievement: A meta-analysis. Psychological Bulletin, 89, 47-62.
- Keller, F. S. (1968). "Good-bye, teacher. . ." Journal of Applied Behavior Analysis, 1, 79-89.
- Kulik, C., Kulik, J., & Bangert-Drowns, R. (1990). Effectiveness of Mastery learning programs: A meta-analysis. Review of Educational Research, 60(2), 265-299.
- Levitz, R. & Noel, L. (2000). The earth-shaking but quiet revolution in retention management. Retrieved July 8, 2002 from <http://www.noellevitz.com/pdfs/Revolution.pdf>
- Martin, D.C. & Arendale, D.R.

Supplemental Instruction (SI) involves pairing high risk courses with small-group sessions in which a tutor experienced with the course helps students with content and study strategies (Martin & Arendale, 1994). Our theory suggests this model is effective precisely because the courses typically used are highly demanding and because small group sessions under the lead of a concerned and helpful tutor are responsive to individual student needs. Faculty and SI tutors need training in how to create an environment where students can experiment with new strategies, get feedback on their effectiveness, and choose the ones that are most effective.

Collaborative techniques such as cooperative learning (Johnson, Maruyama, Johnson, Nelson, & Skon, 1981) have students work together to learn or apply course material. Our theory suggests that the cooperative learning method is effective largely because it is highly responsive to students. But cooperative groups will not work unless instructors learn how to structure them effectively so that students “buy in” to the group process.

In our own classroom we have been working with Keller’s (1968) mastery learning model known as the personalized system of instruction (PSI). A great deal of research supports the validity of this technique (Kulik, Kulik, & Bangert-Drowns, 1990) and it is highly recommended for developmental students (Bonham, 1990). We (Brothen & Wambach, 2000) have described how our application of PSI has enabled students to take control of their learning, develop a sense of self-efficacy, acquire good study habits and skills, and persist until successful. But PSI has foundered on the great deal of organizational work demanded of instructors (Menges, 1994). If developmental educators are to adopt methods with demonstrated effectiveness such as this one, there must be more opportunities for them to get interesting and useful training.

Conclusion

In proposing our theory we recognized that creating demanding and responsive environments would not be a simple task. It takes a great deal of expertise and time to develop assignments at the appropriate level, provide adequate feedback, and respond to students as individuals. Nevertheless, we hoped that our efforts would be helpful to developmental educators. Creating a structure that can help organize our field in practical and effective ways is critical at this point in the history of developmental education. Because developmental education is a young discipline, part of the process of its maturation will involve struggling with competing theories. This struggle should result in research that improves our enterprise. Kurt Lewin’s dictum, that nothing is so practical as a good theory, is worth restating. If we can all apply this to developmental education, we will have taken another step toward meeting our responsibilities as educators and fulfilling the aspirations of our students for a better life.

teachers, assign various papers, and discuss anticipated problems and solutions.

3. Plan a pilot study. Think small, beginning with, perhaps, one or two classes. Consider keeping a log with daily entries to record thoughts, feelings, and insights.
4. Define the role of adjunct faculty. What can they do; what can’t they do? Will they be compensated for extra time requirements? How much time will you be asking them to commit? How will you ensure they share the same understanding of writing that others do?
5. Plan for two types of rater training, one aimed at understanding basic categories and one aimed at rating papers in terms of those categories. The understanding of good writing and the categories by which it is judged are not easy to articulate. The basic categories connect to portfolio purpose and the literature search.
6. Identify the categories before you begin the project, and engage in a small pilot to test their effectiveness. This will avoid changing categories during a project which creates confusion and results in lowered reliability.
7. Provide ample time for discussion and training regarding the meaning of key terms. If the TRRs are not clear, students will not be either.
8. Procedures, as well as categories, need to be thought through and clarified. Students have a right to consistent procedures.
9. Plan to avoid rater overload. If one theme runs through the portfolio literature, it is portfolio assessment results in increased workload. TRR burnout leads to differences in ratings which are a prime source of unreliability. This is another reason for a small pilot to see what is actually involved and how many TRRs need to be involved.
10. Design a simple portfolio project; a complex portfolio is less reliably rated than a simple portfolio. Nystrand, Cohen, & Dowling (1993) found their first study resulted in low reliability. They report increased reliability with a number of changes in the second portfolio assessment project. First, they reduced the number of papers in the portfolio. Then they changed the assessment sequence. Instead of rating all the writing, portfolio by portfolio, they had raters evaluate only texts written in response to a certain prompt. In addition, when raters moved on to another writing assignment, TRRs stopped and recalibrated.

11. Prepare students for a new form of assessment to ensure understanding, comfort, and optimal performance.

Implications: One Portfolio Assessment Project

Using the guidelines discussed above, an excellent opportunity for reliable and valid portfolio assessment is available to individual course instructors. As mentioned previously, a single teacher avoids many of the problems associated with low reliability because she does not have to assess a paper and then have that paper assessed by another rater. By grading all the papers herself, an instructor maintains a higher level of internal consistency and thus higher reliability. The purposes of a portfolio assessment project can include: (a) learning how a portfolio project works, (b) conducting a small pilot to see if portfolio assessment is helpful to one's students and teaching, and (c) examining the role revision plays in writing by being able to compare those students who do multiple revisions with those who do not. The categories of evaluation, or criteria, should be established and clearly explained to students. For instance, an instructor may espouse a formalist philosophy (about 20%), which stresses correct spelling, grammar, and punctuation, and a mimetic philosophy (about 80%), which stresses a clear connection between good writing and good thinking (Fulkerson, 2000). Specifically, then, the persuasive writing would be evaluated on the following criteria: (a) a clear statement of the topic sentence, (b) use of transitional markers, (c) relevant reasons, and (d) support or evidence provided for reasons. Finally, one could collect prewriting and post writing to judge gross improvement. Prewriting can be obtained by asking students to write a persuasive essay before formal instruction begins, and post writing can be collected on the same genre using a very similar topic. At the end of the semester, both the instructor and students could evaluate whether the portfolio model fulfilled the established purpose and goals.

Conclusion

From the traditional educational measurement point of view, portfolio assessment, by and large, has not been reliable, and thus not valid, at least for large-scale high stakes purposes. However, it would be a mistake to think that past conceptions of validity and reliability are static. Validation study is an ongoing field, and there are calls for new conceptions of reliability and validity for many of the new forms of assessment such as portfolio assessment (Hambleton & Sireci, 1997; Moss, 1992).

In our rush to throw out NRTs, we may be creating a vacuum in which we want to nominate portfolios to take their place because portfolio assessment is intuitively understandable by teachers and reflects their teaching values. Maybe our problem is that we operate out of an either-or mentality. Either we

for developing these skills within our courses. One of the reasons we made use of this training was that our faculty agreed that we were all going to take responsibility for developing students' academic skills. We were persuaded by Mike Rose's (1985) argument that literacy was not developed in one composition course, but developed throughout the college experience (Wambach & Brothen, 2002). We all need to play a part in the process. College leaders need to encourage all faculty members to play a role in the skill development process and to use the expertise available in their own faculty to make it happen. Administrators also need to support interventions such as paired courses and supplemental instruction as on-going, not just experimental, parts of the curriculum.

Training

For developmental education to become a national enterprise with recognized techniques and standards we need more agreement on what we do and how we do it. We need to apply the 1995 National Association for Developmental Education (NADE) definition of developmental education as "practice and research within higher education with a theoretical foundation in developmental psychology and learning theory" and recognize the population served as "postsecondary learners at all levels of the learning continuum". This definition allows developmental educators to bring any form of academic support under their wing (cf. Higbee, 1996). It is also consistent with the new ethic that defines retention as everyone in the institution's responsibility and assigns all faculty responsibility for student development. The new definition includes pre-college level courses, learning assistance center activities, supplemental instruction, and freshman seminars, but also allows for the idea that any college course could be developmental education if the instructor is developing students' academic skills as well as their content knowledge.

We have become convinced of the need for a training initiative in developmental education. To illustrate, Grubb (1999) does not advocate business as usual for underprepared students. He strongly advocates teacher training in the community colleges and points out that "developmental education is one of the most difficult teaching challenges and needs to be rescued from its second class status" (p. 174). He also argues that to be successful, developmental education should be "integrated with academic and occupational subjects" (p. 205). Integrating developmental education with the college level curriculum requires an institutional effort and faculty, advisors, and other staff members need training in how to do it.

There are numerous instructional models described by Boylan (1999) and others that are effective alternatives to traditional remedial education. Our theory outlined why and how they might work with developmental students.

instructors provide more rather than fewer testing opportunities and feedback is available quickly. But support does not mean reducing reasonable course demands because students are whining or granting policy exceptions to teary eyed or hostile individuals.

Realistic Expectations for Success

Most college faculty members know how to create demanding courses. However, when courses are demanding, even if they are supportive, some students will fail. Any demanding course requires significant student effort and students who are unwilling to expend effort will fail it. Faculty members cannot teach demanding courses without support from administrators and the college community. If the institution's department chairs and deans believe that retention is the college's most important goal and that disgruntled students must be appeased, then faculty will be unlikely to set high expectations for student performance. If student evaluations of teaching hold too much weight in decisions about faculty members' careers, then teachers will cater to student opinion by making their courses entertaining, reducing expectations, and inflating grades (see Richardson, Fiske & Okun, 1983, and Grubb, 1999 for descriptions of this process). Demanding, supportive courses are designed to prepare students for continued academic success; they are not designed to increase student retention. While some retention specialists such as Levitz and Noel (2000) argue that high institutional standards indirectly promote retention, retention is affected by many factors that are beyond the influence of faculty and staff. For example, our research into reasons why students fail to complete our course successfully shows that almost all students who make an effort but fail have personal and financial problems that affect their educational decisions (Brothen & Wambach, 1999). Also, we find that most students who fail are those who have completed little or no work in our course. If students refuse to engage in the educational process we cannot influence their academic development.

It is sometimes difficult for faculty members who teach introductory liberal arts courses to create assignments that help students develop their reading, writing, and study skills. Graduate training in most fields does not include instruction in how to do this and colleges often attempt to fill the gap by providing faculty with professional development training in pedagogy. Many developmental education faculty trained to teach reading, study skills, and writing know how to create assignments that facilitate the development of critical skills. They have the expertise to train their colleagues from other disciplines in ways to develop effective assignments and give meaningful feedback. We were fortunate to have opportunities to teach paired courses with colleagues who were reading and writing teachers. These experiences had a highly positive impact on our teaching. We also were able to attend workshops on reading and writing across the curriculum, which gave us many useful ideas

ask NRTs to be the sole high stakes decision maker, or we dismiss NRTs from that task altogether and seek another solution. The problem may be our desire to have a single form of assessment to answer all our questions, rather than use a variety of assessment information and make our decisions based on that mix.

The traditional understanding of validity and reliability may not be applicable to portfolio assessment. Perhaps portfolios should not be used for summative evaluation such as making pass or fail judgments but should be used instead for formative evaluation by helping TRRs and students assess development.

There are several possible directions for using portfolio assessment. One is to use it only in single teacher classrooms because there would not be competing norms. Another is to base portfolio assessment on more qualitative understandings of reliability and validity as Pitts, et al. (2001) and Moss, et al. (1992) suggest. Yet another is to recognize that portfolios may be a very sensitive instrument and more useful in revealing developmental learning than in making final judgments. Either way, we can rejoice that we have another instrument in our assessment kit that makes our evaluation more subtle and embrace more traditional forms of testing such as NRTs and CRTs for what they can do, rather than reject them for what they cannot do.

Dale Griffie is an instructor in the English Department at South Plains College, Leelland, Texas.

References

- Adams, T. L. (1995). A paradigm for portfolio assessment in teacher education. *Education*, 115(4), 568-570.
- Arter, J. A., & Spandel, V. (1992). Using portfolios of student work in instruction and assessment. *Educational Measurement: Issues and Practice*, 11, 36-44.
- Bachman, L. F. (1989). The development and use of criterion-referenced tests of language ability in language program evaluation. In R. K. Johnson (Ed.), *The second language curriculum* (pp. 242-258). Cambridge: Cambridge University Press.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford University Press.
- Baumflek, S., Bloom, D., Dettmer, K., DiToro, D., Friedland, A., Gooden, O., Gooding, K., Richardson, A., Hill, J., McQuillan, G., Mlynarczyk, R., Percaccio, F., & Scordaras, M. (1997). Part-timers, full-timers, and portfolio assessment. *Teaching English in the Two Year College*, 25, 308-316.
- Brown, J. D. (1995). Language program evaluations: Decisions, problems and solutions. *Annual*

- Review of Applied Linguistics, 15, 227-248.
- Christian, B. (1993). Freshman composition portfolio in a small college. Teaching English in the Two Year College, 21, 289-2297.
- Cronbach, L. J. (1990). Essentials of Psychological Testing (5th ed.). New York: Harper Collins.
- Darling-Hammond, L. (1994). Performance-based assessment and educational equity. Harvard Educational Review, 64(1), 5-30.
- Erickson, F. (1986). Qualitative methods in research on teaching. In M. Wittrock (Ed.), Handbook on research and teaching (pp. 119-161). New York: MacMillan.
- Falvey, P., & Cheng, L. (2000). Paradigm shift: Teachers' classroom assessment practices in Hong Kong and the PRC. Melbourne Papers in Language Testing, 9(1), 29-50.
- Feldt, L. S., & Brennan, R. L. (1993). Reliability. In R. L. Linn (Ed.), Educational measurement (3rd ed.) (pp. 105-146). American Council on Education. Phoenix, AZ: Oryx Press.
- Fulkerson, R. (2000). Four philosophies of composition. In E. Corbett, N. Myers, & G. Tate (Eds.), The writing teacher's sourcebook (4th ed) (pp. 3-8). New York: Oxford University Press.
- Hambleton, R. K., & Sireci, S. (1997). Future directions for norm-referenced and criterion-referenced achievement testing. International Journal of Educational Research, 27(5), 379-393.
- Hamp-Lyons, L., & Condon, W. (1993). Questioning assumptions about portfolio-based assessment. College Composition and Communication, 44 (2), 176-190.
- Hamp-Lyons, L., & Condon, W. (2000). Assessing the portfolio: Principles for practice, theory, and research. Cresskill, New Jersey: Hampton Press.
- Harrison, S. (1995). Portfolios across the curriculum. WPA (Writing Program Administration), 19 (1/2), 38-48).
- Koretz, D., Stecher, B., Klein, S., & McCaffery, D. (1994). The Vermont portfolio assessment program: Findings and implications. Educational Measurement: Issues and practices, 13 (3), 5-16.
- Messick, S. (1993). Validity. In R. L. Linn (Ed.), Educational measurement (3rd ed.) (pp. 13-103). American Council on Education. Phoenix, AZ: Oryx Press.
- Metzger, E., & Bryant, L. (1993). Portfolio assessment: Pedagogy, power, and the student. Teaching English in the Two Year College, 21, 279-288.
- Moss, P. A. (1992). Shifting conceptions in validity in educational measurement. Review of educational research, 62(3), 229-258.
- Moss, P., Beck, J. S., Ebbs, C., Matson, B., Muchmore, J., Steele, D., Taylor, C., & Herter, R. (1992). Portfolios, accountability, and an interpretive approach to validity. Educational Measurement, 11(3), 12-21.
- Nystrand, M., Cohen, A. S., &

especially true for students who were not in high school advanced placement or accelerated courses that mimic the college environment. Making the transition to autonomous, self-directed learning does not occur overnight. With experience, students acquire knowledge of how to do the usual academic tasks, how much time these tasks require, and how to build that time into their schedules. If the instructors who teach freshman courses make them too easy, students do not learn the metacognitive skills they will need to become successful in demanding training programs or careers. If the courses are too difficult, many students will fail no matter how much effort they expend. We believe that all freshman courses should be designed so that everyone has to exert significant effort to pass and that all who make this effort will pass.

Demanding Courses

Demanding courses set high but attainable standards for mastery of content and skills. In most disciplines there is at least some consensus on what content should be included in beginning courses and what skills should be mastered first. For example, in our discipline of psychology there is great similarity in the content of all major introductory textbooks. There is also the expectation that students should be able to answer multiple-choice items, which require recognition memory, and to provide brief answers to literal questions. In our demanding general psychology course, we expect students to read and understand the entire textbook. Our tests require remembering new vocabulary, important theories, the relationships between concepts, results of research, knowledge of details, and application of information to new situations. With experience, instructors develop a feel for the mix of test items that distinguish students who have worked hard from those who have not. Challenge is not to be confused with poorly worded assignments and test questions that even the hardest working students cannot figure out.

Supportive Courses

Supportive courses provide useful feedback to students about their progress in attaining the course goals. Students have many opportunities to demonstrate their knowledge and supportive professors treat all students with respect. Respect includes trying to get to know students as individuals, showing them that you are aware of and concerned about their performance, and fully informing students of all course policies. In our general psychology course, students receive a detailed syllabus with course objectives, grading policies, detailed instructions for completing assignments, and dates when assignments are due and exams are scheduled. We try to anticipate what content will be controversial and select textbooks and prepare materials that address those controversies. In supportive courses, students are given the opportunity to express their opinions about issues that are of interest to them. Supportive

freshman seminars. However, if the only option available to an institution is to offer students isolated pre-college level skills courses, then these courses must be challenging and clearly connected to the rest of the curriculum.

We also asserted that the developmental education environment must be responsive. To facilitate this we suggested several good educational practices that might be particularly important for developmental education. For example, courses should be small enough to allow students and teachers to get to know one another. Also important are instructors, advisors, and other staff with good listening and conflict resolution skills, respect for cultural differences, and an ability to communicate their expectations to students through well-crafted assignments, detailed syllabi, and coherent services.

A key element in creating a responsive developmental education environment is timely and useful feedback. Highly skilled students have the ability to provide their own feedback and know whether they are doing well in their courses. They know the difference between simply doing and actually learning from assignments. Many first year students, especially those who were not high achievers in high school, still need to acquire these skills. To do so they require regular feedback that identifies how much they have learned and where they should focus additional practice. We challenged developmental educators to find ways to deliver helpful feedback to their students.

The concepts of self regulation, demandingness, and responsiveness explain quite well how developmental students came to be as they are and what may be required to help them become successful. These concepts can also be used more generally in thinking about institutional structures, educational practices, and teacher behaviors.

Developmental Theory & Practical Tasks

The child development literature shows that the most successful parents are demanding and supportive and expect their children to display age appropriate maturity in their conduct. These expectations exist in the context of a respectful relationship between the parent and child. In this relationship, parents express warmth and caring, explain expectations and rules, and avoid punishment as much as possible. Specifying appropriate expectations for traditional age college freshmen is also critical.

New college students anticipate more freedom in college courses, but many students, no matter what their high school achievements were, are not equipped to achieve in environments that are high in demand and low in support (e.g., the prototypic lecture class with a mid-quarter and a final). Also, high school experience trains students to rely heavily on teachers to tell them exactly what will be on tests and to allow plenty of in-class time to do assignments. This is

Dowling, N. M. (1993). Addressing reliability problems in the portfolio assessment of college writing. Educational Assessment, 1(1), 53-70.

Padilla, A., Aninao, J., & Sung, H. (1996). Developing and implementation of student portfolios in foreign language programs. Foreign Language Annals, 29(3), 429-438.

Pitts, J., Coles, C., & Thomas, P. (2001). Enhancing reliability in portfolio assessment: Shaping the portfolio. Medical Teacher, 23(4), 351-356.

Seliger, H. W., & Shohamy, E. (1989). Second language research methods. Oxford: Oxford University Press.

BETTER TUTORING FOR NON-NATIVE SPEAKERS OF ENGLISH IN COLLEGE TUTORING AND WRITING CENTERS

By Steven Bookman, Lehman College, City University of New York

Abstract

Most tutors in college and university tutoring and writing centers do not have a background in linguistics and second language acquisition. This paper proposes the need for a specially trained group of “ESL literacy tutors” to work with college students who are non-native speakers of English. If literacy tutors had a sufficient background in linguistics and second language acquisition, they would be better equipped to handle non-native speakers’ requests for assistance.

Background

There is an increasing number of non-native speakers of English from all language backgrounds in universities and colleges in the United States. Many of them seek tutoring in pronunciation, in English as a second language (ESL), and in reading and writing. Many tutoring and writing centers in colleges and universities, where there is a large non-native speaker population, do not have enough tutors, knowledgeable and comfortable in teaching English language skills, to effectively meet the demands of English language learners (Ronesi, 1995; Kinkead & Harris, 2000). The purpose of this paper is to propose the need for a specially trained group of “ESL literacy tutors” to work as peer or writing center tutors alongside other tutors. These ESL literacy tutors would take a training class in second language acquisition and linguistics to help them cope more effectively with English language learners struggling with the high level of English necessary for post-secondary studies.

An ESL literacy tutor would help students with English as a second language issues including pronunciation, editing, study skills, and academic reading and writing skills. Such a tutor would guide students to become proficient in studying techniques, reading, and writing, and to become literate in English. This ESL literacy tutor should have a basic understanding of phonology (sound system), morphology (word parts), syntax (grammar), semantics (meaning), and pragmatics (using language socially) of learners’ native languages in order to more effectively help them learn English. In other words, the tutor should understand a concept well enough to explain it in simple terms, but does not have to be an expert. Also, all non-native speakers, regardless of their native



JOIN THE CONVERSATION

DEVELOPMENTAL THEORY: THE NEXT STEPS

By Thomas Brothen and Cathrine Wambach, General College, University of Minnesota

Recently, we (Wambach, Brothen, & Dikel, 2000) proposed a comprehensive theory that focused on the developmental education environment. We set our theoretical foundation on the solid body of research from developmental psychology that examines developmental outcomes, research on schools as caring communities, and students’ adjustment to college. We united three basic concepts in our perspective: demandingness, responsiveness, and self-regulation and asserted that these concepts could be used to organize, explain, and predict useful techniques for practitioners. Our goals were to improve practice and stimulate further research on developmental education practices that facilitate student success.

Self-regulation is central to our theoretical perspective because much of what students should have learned by the time they reach post secondary education involves extra-scholastic skills and abilities and a level of independence and maturity that will allow them to follow a self-directed path through graduation. In developmental education programs, students, still developing self-regulatory skills, not only learn course material, they acquire a repertoire of skills that will help them succeed academically throughout and beyond their time in college. It is this emphasis on personal development, together with academic training, that separates developmental education from the rest of higher education.

We reviewed a body of research showing clearly that self-regulation is developed in demanding and supportive environments. Therefore, our first assertion was that developmental educators must be demanding. For this to become a reality, we argued that the curriculum should be content-based and worth credit toward degrees. Currently, much of developmental education consists primarily of stand-alone, non college-credit skill development courses in reading, writing, and mathematics. Students often resist these courses because they overestimate their level of academic skills and see no reason they should wait to begin their regular studies. We recommended Boylan’s (1999) options for merging developmental education with an institution’s degree credit curriculum through devices such as supplemental instruction and

References

- Arendale, D. (2001). The supplemental instruction supervisors manual. Kansas City, MO: University of Missouri-Kansas City.
- Blanc, R., DeBuhr, L., & Martin, D. (1983). Breaking the attrition cycle: The effects of supplemental instruction on undergraduate performance and attrition. Journal of Higher Education, 54, 80-89.
- Blumenfeld, P. C. (1992). Classroom learning and motivation: Classifying and expanding theory. Journal of Educational Psychology, 84(3), 272-281.
- Eccles, J. S., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., & Midgely, C. (1983). Expectancies, values, and academic behaviors. In J. T. Spence (Ed.), Achievement and Academic Motivation (pp. 75-146). San Francisco, CA: W. H. Freeman.
- Kenney, P. (1989). Effects of Supplemental Instruction on student performance in a college level mathematics course, University of Texas-Austin. Dissertation Abstracts for Social Sciences, 50, DA8909688.
- Kochenour, E. O., Jolley, D. S., Kaup, J. G., Patrick, D. L., Roach, K. D., & Wenzler, L. A. (1997). Supplemental instruction: An effective component of student affairs programming. Journal of College Student Development, 38(6) 577-586.
- Longo, D. A., Lent, R. W., & Brown, S. D. (1992). Social cognitive variables in the prediction of client motivation and attrition. Journal of Counseling Psychology, 39(4) 447-452.
- Petri, H. L. (1995). Motivation: Theory, research, and applications. Pacific Grove, CA: Brooks/Cole Publishing Company.
- Pintrich, P. R., & Schrauben, B. (1992). Students' motivational beliefs and their cognitive engagement in classroom academic tasks. In D. H. Schunk & J. L. Meece (Eds.), Student perceptions in the classroom: Causes and consequences (pp. 149-183). Hillsdale NJ: Lawrence Erlbaum Associates.
- Weiner, B. (1986). An attributional theory of motivation and emotion. New York: Springer-Verlag.
- Wolters, C. A. (1998). Self-regulated learning and college students' regulation of motivation. Journal of Educational Psychology, 90(2), 24-235.

language, have different experiences. The literacy tutor must realize this and treat each student as an individual and not as part of a homogenous group. This paper is written based on the author's experiences of tutoring English as a second language, observing ESL tutoring sessions, and talking with English language learners, who were being tutored by other tutors, on several college campuses.

Essential Linguistics Knowledge

Phonology

Today's tutoring centers on diverse college campuses should include literacy tutors who are knowledgeable about different grammatical and phonological systems, thus having an idea of what types of pronunciation errors their students may make. If ESL students are too shy to speak because of difficulties pronouncing English, they may not fully understand the material because they will not ask questions in class. Being able to work with a tutor in pronunciation can be very important for their academic development.

To work with an English language learner in pronunciation, the literacy tutor must have some knowledge of the English sound system and should be familiar with the most common errors of the students' languages encountered in the tutoring center. Some languages, such as Spanish, have a five-vowel system (Stockwell & Bowen, 1965); because of this, English vowels are difficult for Spanish speakers to pronounce because English has a twelve-vowel system (Ladefoged, 1999). Spanish has a one-to-one correspondence between a vowel and its sound, as opposed to English. Also, some languages, such as Korean, have rounded and unrounded vowel counterparts (Lee, 1999; Lee & Ramsey, 2000). English, on the other hand, does not have any rounded and unrounded vowel counterparts (Ladefoged, 1999). As a result, English language learners will often pronounce the target vowels too high or too low.

The ESL literacy tutor also needs to know why non-native speakers pronounce consonants a certain way. When Koreans and Japanese pronounce an /r/, it sounds like an alveolar sound which is between an /r/ and /l/ or the /r/ or /l/ may be used interchangeably (Cheng, 1993) because they do not have an /r/ in their phonetic systems (Bell, 1996; Ladefoged, 1999). As a result, Japanese, and to an extent Korean, English language learners may say 'read' when they mean 'lead'. Similarly, Koreans will perceive 'zoo' as 'Jew' (Sohn, 1999).

Syntax

In addition to having a basic knowledge of phonology, knowing how to explain different parts of speech and their word order in sentences for English and the

most common languages spoken on campus is necessary. Korean and Japanese, for example, have similar grammars. They have a subject-object-verb word order (Cheng, 1993; Sohn, 1999; Lee & Ramsey, 2000). The subject can move around as long as it goes before the predicate and verb, and the predicate contains the verb or adjective (Sohn, 1999). They contrast with English, which has a subject-verb-object word order. It is also important to note that the word order for questions varies from language to language. In Korean, a question has the same word order as a statement; in other words, there is no inversion of the subject and verb in questions, as it occurs in English (Cheng, 1993; Sohn, 1999).

Many languages do not use pronouns or plurals. For example, in Japanese, personal pronouns are omitted since they are inferred from the context (Cheng, 1993). Also, the Chinese languages, Korean, and Japanese do not use any articles or plural markers (Cheng, 1993; Shim, 1999; Sohn, 1999; Lee & Ramsey, 2000). Subject-verb agreement in Tagalog is not inflected, meaning that singular and plural forms are the same (Cheng, 1993). On the other hand, there are languages, such as Spanish, that pluralize adjectives and use articles in the majority of cases where a noun is used (Kayser, 1993).

Pedagogy

The ESL literacy tutor should be conscious of the disparity between ways of teaching and responding in the English language learner's native country and the United States (Cheng, 1993). Having knowledge of teaching methods from other countries and of different cultures is important when tutoring non-native speakers in order to understand why the student may respond differently than an American student. In some cultures, for instance, memorization is used as the preferred method of teaching. As an example, Koreans are taught to memorize words and phrases in high school (Bell, 1996); if they have not learned such words or phrases, they may not use them. In some cultures, students are taught not to volunteer information (Bell, 1996). In fact, teachers are not to be interrupted. In the United States, on the other hand, the students often must participate and are evaluated on their participation. Silence is an important element in communication.

In sum, the literacy tutor needs to have a basic understanding of linguistics and second language acquisition, as well as knowledge of the characteristics of different languages and cultures. This knowledge will help the literacy tutor understand why the student makes certain types of errors and, thus, how to approach the material.

unable to attend sessions. This difference is approximately a quarter of a letter grade even at lower levels of SI session attendance ($2.76 - 2.48 = 0.28$). Therefore, actual participation in the SI sessions was shown to provide grade benefits beyond the benefits due to the increased general motivation of SI participants. Also, more frequent participation in SI sessions was shown to be more beneficial than less frequent participation. Increased participation resulted in a grade increase of 0.18 ($2.94 - 2.76$) grade points.

The grade increases associated with SI session participation are believed to be due to enhanced learning resulting from the leaders' use of the techniques in the SI model. The sessions feature safe environments for students to ask questions and opportunities for guided practice and interaction with peers. The educational benefits of guided practice, support, and interaction are well accepted by education professionals. At the college level, these are features that are often missing unless provided by an academic support program such as Supplemental Instruction.

Conclusion

This study showed that participation in Supplemental Instruction sessions is associated with higher grades in a quantitative college chemistry course. It also revealed that increased motivation by itself is associated with higher grades presumably due to higher levels of effort expended. In this study, these higher levels of effort are inferred, not documented. Future studies should be conducted to determine the amount of productive efforts outside of class that students need to achieve academic success. To obtain a comprehensive understanding of the causes of student success, the variables of academic preparation, motivation, and actual effort should be examined. In looking at general issues of academic preparation and success at the university level, these studies could begin to determine how much effort outside the classroom is necessary for success and approximately how much extra work might be required to make up deficiencies in aptitude.

Kenneth W. Gattis, Ph.D., is the Director of the Undergraduate Tutorial Center at North Carolina State University in Raleigh, North Carolina.

Analysis of Covariance (ANCOVA) is a statistical technique that can also remove the effects of unwanted influential independent variables in order to determine if significant differences can be found due to other independent variables. The variable removed from consideration when performing an ANCOVA is called a covariate. An ANCOVA performed on the SI data (see Table 4) showed that the five groups of students had significantly different average Chemistry II grades ($p = .0177$) -adjusted for the covariate AI. This result is confirmed visually by looking at the adjusted course grades shown in Figure 1.

Table 4. Analysis of covariance for dependent variable Chemistry II grade.

<u>Source</u>	<u>df</u>	<u>Sum Squares</u>	<u>Mean Square</u>	<u>F-value</u>	<u>p-value</u>
Model	5	72.20	14.41	11.04	0.0001
Error	109	142.54	1.31	1.31	
Corrected total	114	214.74			
<u>Model Sources</u>					
AI	1	55.83	55.83	42.69	0.0001
Group	4	16.37	4.09	3.13	0.0177

Discussion

This motivational control study affirms the reality of potentially higher academic performance for groups of students that voluntarily participate in academic support programs. Research question one can be answered in the affirmative. The significant increase in course grades of the motivational control group over the nonparticipating group shows that student grade increases in college chemistry are associated with students' reported motivation to attend SI sessions. After removing the effects of academic preparation (as estimated by AI), the only known difference between these two groups was the motivation to attend SI as expressed on the survey.

The analysis showed motivation to be an important factor in grade performance whether students used SI or not. This is the factor cited when skeptics say that "top students" will perform well regardless of the quality of instruction. The motivational control group created by the survey responses isolated some of the highly motivated students and quantified the course grade benefits of their increased motivation. This difference appears to be close to half a letter grade ($2.48 - 2.04 = 0.44$).

The study also bolstered claims of improved performance for students participating in Supplemental Instruction sessions. Research question two can be answered in the affirmative. Students who participated in SI sessions earned higher Chemistry II grades than students who were motivated yet

Related Areas

Contrastive Analysis

Gaining knowledge about the characteristics of the non-native speaker's first language is helpful because the literacy tutor can have a better idea as to what types of errors could be made. Thus, by predicting the errors made by the English language learner, the tutor can modify the student's accuracy. This theory, called Contrastive Analysis Hypothesis, is a behaviorist approach, which emphasizes that all errors are to be eliminated (Brown, 1994; Benson, 2002; Haley & Rentz, 2002). Although the goal of elimination of all errors is considered unrealistic today, it can be useful for literacy tutors to have this knowledge of the characteristics of the major languages on their diverse campuses. This can promote better tutoring because the ESL tutors can cater to the student population on their campuses. One way to implement this is by having a databank in the tutoring center where tutors can access this information anytime.

Natural Order of Acquisition of Morphemes

Roger Brown (1973) traced the order of acquisition of fourteen morphemes in children whose native language was English. Brown's study sparked interest in second language acquisition researchers studying the order of grammatical morphemes in ESL children learning English. The order is almost identical when compared to children who are native English speakers (Dulay & Burt, 1974). In that adult second language acquisition follows the same pattern of acquiring the same order of morphemes as children (Bailey, Madden, & Krashen, 1974; Larsen-Freeman, 1976; Krashen, 1987), it appears that every English language learner goes through, generally, the same specific order of learning morphemes. This knowledge is important because it tells the literacy tutor which morphemes (such as plural -s, -ed, and -er) are acquired early and late and, thus, indicates a hierarchical order in which errors should be addressed.

Interlanguage Development

Having knowledge of the stages of second language acquisition is crucial because Pienemann (as cited in Haley & Rentz, 2002) states "it is important to know what is learnable at what point in time" (p. 1). There are four stages every English language learner goes through. Throughout these stages, the student will create a temporary language until the target language is mastered. This temporary language is called an interlanguage (Brown, 1994). In the beginning (stage 1), the ESL student is vaguely aware that there is some systematic order to a particular class of items. The second language learner

makes wild, inaccurate guesses or random errors (transfer errors), which have the characteristics from the student's native language (Brown, 1994). Then, the learner begins to separate the two languages and to internalize rules in the emergent state (stage 2). The student makes developmental errors or intralingual errors, meaning the errors do not come from the first language; they come directly from the target language (Brown).

In the stabilization stage (stage 3), the learner commits fewer errors in his target language and is gaining fluency (Brown, 1994). The second language learner self-corrects, and the use of the second language is close to that of the native speaker. Finally, in the post systematic stage (stage 4), the English language learner becomes fluent in the new language. The student has mastered both language systems and keeps them separate (Brown).

By knowing the stages of language development, the literacy tutor can see where the non-native speaker is in his language learning. Knowing this, the tutor will be able to make a decision of where and how to begin language assistance.

Tutoring Procedure

Errors should be addressed through a hierarchical teaching method according to the purpose of the assignment, the type of error (Ferris, 2002), and the importance of the error. Content comes before grammar except in the cases where grammar is being tested. Errors may also mean that the second language learner may not have the vocabulary or language skills to express himself well enough to be understood. ESL literacy tutors need to know how to assess non-native speakers' proficiency when working with them. Corder (as cited in Brown, 1994) proposed a procedure to identify errors in non-native speakers. First, the evaluator, who is the literacy tutor for the purposes of this paper, checks that the sentence in the target language makes sense in the context. If the sentence does not contain the appropriate syntax, he needs to find out if the English language learner translated this sentence into his native tongue first and what his first language is. The final step is to reconstruct the sentence into the target language, explaining how to use it correctly.

When literacy tutors work with second language learners, it should be with context-embedded text (Bell, 1996); in other words, the grammatic structure being addressed should be in a text. When reviewing grammatical concepts, the literacy tutor should present the concept in English, and, then, the same concept can be shown in the tutee's first language. This is one way to get the student to see the what and the why of what is wrong (Nakomori, 2002). For example, relative clauses are very difficult for native Japanese speakers learning English because there are no relative clauses in Japanese (Nakomori). The literacy tutor can explain to the non-native speaker that relative clauses are

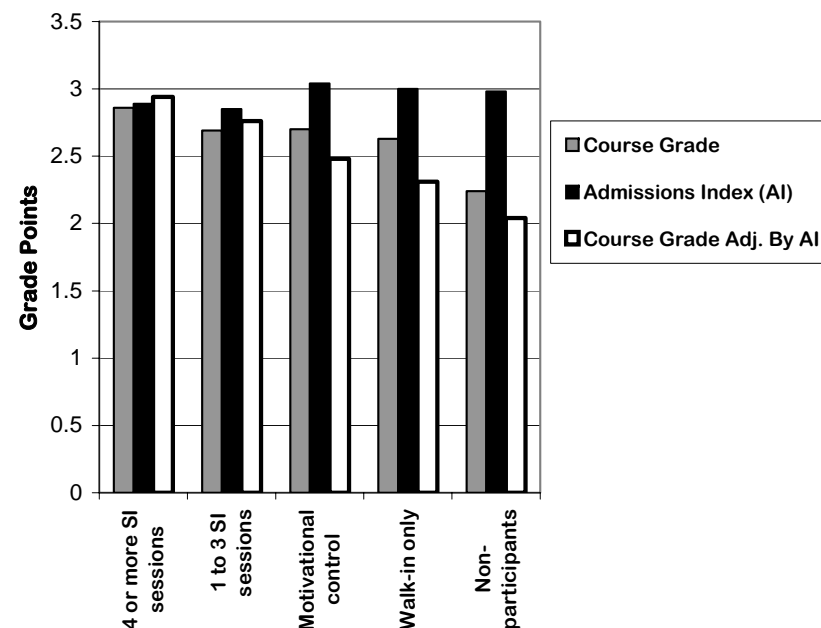
The Chemistry II grades were adjusted for the group differences in average AI. In this way, the effect of academic preparation was removed from the analysis. Table 3 shows that the highest adjusted course grades were earned by students attending 4 or more SI sessions followed by students attending 1 to 3 sessions. The motivational control group averaged higher than both the non-participants and the walk-in students but lower than both groups of SI participants.

Table 3. Average Chemistry II grade, adjusted by Admissions Index (AI).

Group	Number	Avg. Grade Chem II	Avg. AI	Chem II Grade Adjusted by AI
4 or more SIs	41	2.86	2.89	2.94
1 to 3 SIs	33	2.69	2.85	2.76
Motiv. Control	11	2.70	3.04	2.48
Walk-in Only	9	2.63	3.00	2.31
Non-participants	48	2.24	2.98	2.04
TOTAL	142			

Figure 1 illustrates the effects of the grade adjustments, called least square means. Since both groups of SI participants had lower average AIs, the average course grade for these groups was adjusted upwards slightly. The other three groups had relatively high average AIs, and the average course grade for these groups was adjusted downward.

Figure 1. Effect of adjusting Chemistry II course grade AI



Students who listed at least one available time coinciding with scheduled SI session times were placed in one of three other groups based on the extent of their participation in the SI sessions and/or their use of a chemistry drop-in service. Those students who only used the drop-in service were placed in a separate group. The purpose was not to study the effectiveness of the drop-in service, but to remove the contaminating effect of the use of a separate academic support service on chemistry achievement. Frequent SI participants, defined as those attending four or more sessions, were separated from the less frequent participants. The groupings of those attending one to three sessions and those attending four or more sessions were chosen in order to include approximately equal numbers of students in each group and to represent different levels of effort in SI session participation.

Grade data was collected at semester end from the Registrar's office. Statistical analyses were performed to determine if grade differences existed between groups. Because academic ability and preparation are so important for success in "high risk" courses such as chemistry, it was decided that the Chemistry II grades would be adjusted for the group differences in average Admissions Index (AI). Analysis of covariance (ANCOVA) would be used to remove such differences and determine whether significant differences could be found.

Results

As seen in Table 2, students who participated in four or more SI sessions earned the highest average grades, followed by the students in the motivational control group and the group of students attending one to three SI sessions.

Table 2. Chemistry II grade and Admissions Index (AI) by group.

Group	Chemistry II Grade		Admissions Index (AI)	
	Mean	Std. Dev.	Mean	Std. Dev.
4 or more SIs	2.86	1.02	2.89	0.247
1 to 3 SIs	2.69	1.23	2.85	0.388
Motiv. Control	2.70	1.25	3.04	0.354
Walk-in Only	2.63	1.68	3.00	0.382
Non-participants	2.24	1.60	2.98	0.395

The five groups varied in their aptitude and academic preparation, as estimated by Admissions Index (AI). As seen in Table 2, the students in the motivational control group had the highest average AI, followed closely by the Walk-in Only group and the non-participants. Both groups of SI participants had average AIs significantly lower than the other groups.

formed by inserting a relative pronoun and a simple sentence after the head noun (Nakomori). An example in English and Japanese follows below (Nakomori p. 31):

1. The dog [which is running over there],
2. Over there running dog [mukoude hashitteiru] inu

Another method that recent research has suggested to be effective with grammatical problems is *focus on form* (Haley & Rentz, 2002). This method lets the non-native speaker notice a particular grammatical form by focusing only on that form. This method prevents errors from becoming habitual and increases the English language learner's pace of learning structures, allowing the student to notice a particular error more often.

Discussion

This paper discusses how having knowledge of linguistics, second language acquisition, and characteristics of the second language learner's native languages and cultures can help the specially trained ESL literacy tutor better assist the non-native speaker. The author is not suggesting that literacy tutors have to be linguists or English as a Second Language teachers, but they should possess a basic knowledge of second language acquisition and linguistics. If English as a second language teaching methods are used, the ESL literacy tutor can make the tutoring session much more effective for the non-native speaker.

There are two ways that tutors can be trained depending on the situation in the college. One way, as discussed in this paper, is to have a small group of literacy tutors specially trained for handling all types of ESL requests for assistance, including pronunciation. All tutors would be trained in the three most important areas: grammar, teaching methods, and culture.

In schools where this specialization isn't feasible due to size or resources, in-service workshops, conducted by experts from the same or a different university can be held to train general tutors with no background in linguistics or second language acquisition to be better prepared to work with non-native speakers. In these workshops, basic teaching strategies, grammar, language characteristics, and cultural learning models should be emphasized. Also, in every tutoring center, there should be an appendix of resources and suggested readings (i.e. summaries in research) that will build upon the skills the tutors already possess.

Steven Bookman is a Master Tutor at Lehman College, City University of New York, an ESL instructor at Westchester Community College, and a graduate student in Applied Linguistics at Lehman College.

References

- Bailey, N., Madden, C., & Krashen, S. (1974). Is there a "natural sequence" in adult second language learning? *Language Learning*, 21, 235-43.
- Bell, M. A. (1996). *Teaching pronunciation and intonation to E.F.L. learners in Korea*. Retrieved April 16, 2002, from www.geocities.com/Athens/Acropolis/9583/index.html.
- Benson, C. (2002). Transfer/cross-linguistic influence. *ELT Journal*, 56(1), 68-70.
- Brown, H. D. (1994). *Principles of language learning and teaching* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall Regents.
- Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.
- Cheng, L. L. (1993). Asian-American cultures. In Dolores E. Battle, *Communication disorders in multicultural populations* (pp. 38-77). Stoneham, MA: Butterworth-Heinemann.
- Dulay, H. & Burt, M. (1974). Natural sequences in child second language acquisition. *Language Learning*, 24, 37-53.
- Ferris, D. (2002). *Treatment of error in second language student writing*. Ann Arbor, MI: The University of Michigan Press.
- Haley, M. H. & Rentz, P. (2002). Applying SLA research and theory to practice: What can a teacher do? *TESL-EJ*, 5(4). Retrieved April 2, 2002, from <http://www-writing.berkeley.edu/TESL-EJ/ej20/a2.html>.
- Kayser, H. (1993). Hispanic cultures. In D. E. Battle, *Communication disorders in multicultural populations* (pp. 114-157). Stoneham, MA: Butterworth-Heinemann.
- Kinkead, J. & Harris, J. (2000). What's next for writing centers? *Writing Center Journal*, 20(2), 23-24.
- Krashen, S. (1987). *Principles and practice in second language acquisition*. United Kingdom: Prentice-Hall International.
- Ladefoged, P. (1999). American English. In Cambridge University Press (Ed.), *Handbook of the International Phonetic Association. A guide to the user of the International Phonetic Alphabet* (pp. 41-44). Cambridge, UK: Cambridge University Press.
- Larsen-Freeman, D. (1976). An explanation for the morpheme acquisition order of second language learners. *Language Learning*, 26, 125-34.
- Lee, H. B. (1999). *Korean*. In Cambridge University Press
- particular course being studied at the particular point in time.
- Using these measures of students' prior knowledge, efforts in participating in SI sessions, and motivation to attend SI sessions, the following questions were addressed:
1. Are student grade increases in college chemistry associated with students' reported motivation to attend SI sessions?
 2. Will productive effort in SI sessions result in higher average college chemistry grades than students who are motivated yet unable to attend sessions?

Method

The study used two sections of a second introductory quantitative chemistry course taken by many students in the sciences and engineering. The same instructor taught both classes. SI sessions were available to students in both sections but were not a part of the regular class schedule. Some students were unable to attend any of the four weekly SI sessions because of schedule conflicts. A small chemistry walk-in service near the chemistry classroom and operated by SI leaders was also available to students during three or four hours each day. Most students who used this service also participated in SI sessions, and students who only used the walk-in service tended to be infrequent users.

A survey was developed to identify a group of students who were not able to attend Supplemental Instruction sessions and who shared the same motivational state as students who did participate in SI sessions. The survey, completed by students during the first week of classes, was used to classify students by their motivation to attend SI and their availability to attend any of the four weekly sessions. In both sections, the researcher first explained the benefits of attending SI sessions, then explained the purpose of the survey, and finally administered the survey. The survey asked students whether they were interested in participating in SI sessions and, if so, to circle all hours that they were available to attend. The survey was completed before the weekly SI session times were announced. Of the 191 students who completed the course in the two sections studied, 142 of them were present on the day of the survey.

For the purpose of the post-semester analysis, students were grouped based on the survey results and their participation in academic support services during the semester. Those students who by their own choice used no academic support services were placed in the non-participant group. Students who were interested in attending SI sessions, but who had time conflicts with all the sessions, were placed into the motivational control group. These students would be presumed to have the same quality of high motivation (at least, in regard to SI attendance) as those who participated in SI sessions.

performance in SI-supported courses.

Despite the use of Admissions Index as a control, some skeptics will still question the validity of such results. They may argue that even though AI is highly predictive of chemistry course grades, there are still other motivational factors involved. It's possible that once in college, one student will be more motivated to succeed than another student who entered college with exactly the same Admissions Index. Because students have the choice of whether or not to participate in Supplemental Instruction sessions, the possibility of self-selection bias still leaves the question of whether grade increases are due to program effectiveness or student motivation.

Research Questions

There are obviously many factors that affect grades earned by students in college chemistry classes, whether the classes are supported by Supplemental Instruction or not. The biggest general factors would seem to be prior knowledge, effort, and motivation.

Although the best measure of students' prior chemistry knowledge would be a comprehensive chemistry exam, such a measure was not available to the researcher. However, indicators of past general academic performance such as previous grade point average or Admissions Index (AI) can be used as proxy measures for chemistry knowledge.

It is difficult to estimate the total effort exerted by students in studying chemistry at the university. When students have access to Supplemental Instruction, a portion of each student's overall effort can be measured by the number of SI sessions attended. However, the number of hours attending SI sessions by even the most frequent participants represents but a fraction of the total hours that most students need to study for SI-supported (high-risk) courses.

Motivation is important in that it sustains effort, leading to higher total amounts of effort exerted over the course of a semester. Measured motivation can help to account for differences in effort not measured directly (Blumenfeld, 1992; Longo, Lent, & Brown, 1992). Simple surveys can be designed to identify students who are interested in attending SI sessions but are prevented from attending due to time conflicts. These students can be assumed to share the same motivations as SI participants but differ in that they can't benefit from the actual sessions. These motivated non-attending students, who normally would be classified as non-participants in SI program assessments, can be placed in a motivational control group. Although expressed interest in SI participation is only a partial measure of students' general motivation for success, it captures some of their motivation for putting forth effort in the

(Ed.), Handbook of the International Phonetic Association. A guide to the user of the International Phonetic Alphabet (pp. 120-123). Cambridge, UK: Cambridge University Press.

Lee, I. & Ramsey, S. R. (2000). The Korean language. Albany, NY: State University of New York Press.

Nakomori, T. (2002). Teaching relative clauses: How to handle a bitter lemon for Japanese learners and English teachers. ELT Journal, 56(1), 29-40.

Ronesi, L. (1995). Meeting in the writing center: The field of ESL. TESL-EJ, 1(3). Retrieved April 20, 2002, from <http://www-writing.berkeley.edu/TESL-EJ/ej03/al.html>.

Shim, R. J. (1999). Codified Korean English: Process, characteristics, and consequence. World Englishes, 18(2), 247-259.

Sohn, H. (1999). The Korean language. Cambridge, UK: Cambridge University Press.

Stockwell, R. P. & Bowen, J. D. (1965). The sounds of English and Spanish. Chicago: The University of Chicago Press.

RESPONDING TO SELF-SELECTION BIAS IN ASSESSMENTS OF ACADEMIC SUPPORT PROGRAMS: A MOTIVATIONAL CONTROL STUDY OF SUPPLEMENTAL INSTRUCTION

By *Kenneth W. Gattis, North Carolina State University*

Abstract

A motivational control study of students participating in Supplemental Instruction (SI) sessions in college chemistry showed that participants benefit from SI sessions to an extent that cannot be explained by their higher levels of motivation. Motivation is shown to be an important factor in grade performance, whether students use SI or not. Actual SI attendance is shown to provide additional grade benefits. The effectiveness of Supplemental Instruction is thought to be due to enhanced interactivity. SI provides students with a productive hour of learning, featuring a psychologically safe environment for asking questions and opportunities for guided practice.

Introduction

By definition, academic support programs operate outside the bounds of the required, structured components of college courses. Such programs may include services such as one-on-one tutoring, drop-in tutoring centers, and group tutorials such as Supplemental Instruction. Students use these services voluntarily, some occasionally, and others more often. A large number of students do not use any support services. Voluntary participation causes a substantive methodological problem in the statistical analysis of the outcomes of these programs. The validity of positive assessment findings (for example, regarding student grades or retention rates) may be threatened by self-selection bias due to the inclusion of a volunteer sample. By deciding to use a support service, students essentially choose to become a member of what becomes the experimental group in the assessment of program results. The use of this volunteer sample in the data analysis provides an alternative explanation for positive results that would otherwise be attributed to the effectiveness of the program. Because of the extra efforts they display in using an academic support service, these “volunteers” initially may be regarded as having increased motivation to do all the things necessary for success in a college course. If these students are indeed more motivated, assessments involving them may inadvertently detect motivationally caused differences in

southeastern United States. The SI program employs 15 to 20 SI leaders each semester to support the instruction in large lecture section classes. SI sessions are offered to students in most sections of freshman and sophomore chemistry and selected sections of biology, mathematics, computer science, and physics. For each class, three or four SI sessions are offered at various times each week, usually during the late afternoon and early evening. During the semester, students attend as many SI sessions as they wish. While session attendance is recorded, students do not receive any extra credit for attending and are informed of this fact.

Approximately 1500 different students attend the sessions each semester. Many students attend only once or twice, but a number of students make the voluntary SI sessions a regular part of their weekly studies. On average, students attend four to five sessions each semester.

All during the 1990s, statistics showed consistently higher course grades associated with increased SI attendance. Table 1 shows the average grades and numbers of participants for six attendance levels. This data is the average of eight consecutive semesters in the mid-1990s.

Table 1. Average course grades and levels of SI participation

<u>Sessions Attended</u>	<u>Avg. Number of Students</u>	<u>Avg. Admissions Index (AI)</u>	<u>Avg. Grade Pts.* in SI Course</u>
0	2606	2.83	2.15
1-2	723	2.81	2.32
3-5	285	2.79	2.47
6-8	129	2.80	2.62
9-13	100	2.81	2.86
Over 13	60	2.78	3.06

*4 grade points for A, 3 for B, 2 for C, 1 for D, 0 for F.

In studying the relationship between SI attendance and course grades, researchers have access to “Admissions Index” (AI) for most students. AI is a predictor of first year grade point average and is derived from high school grade and admissions test information. Because of its roots in both long-term performance measures (high school grades) and standardized tests, it reflects a combination of aptitude, concept knowledge, and effort. It is also highly correlated with final grade in most courses, including the chemistry course in this study. Using Admissions Index as a control variable can help to answer critics who may contend that only the strongest students attend SI sessions. If the students attending SI sessions were stronger on average, one would expect them to have a higher average AI. However, as seen in Table 1, these students have a slightly lower AI than non-participants yet show better grade

A 1996 study of the relationship between SI participation and final course grades addressed the issue of self-selection bias with the creation of a motivational control group (Arendale, 2001). This group consisted of non-participating students who indicated high motivation to attend sessions on a survey taken on the first day of class and attributed their lack of SI session attendance to class or work conflicts on a survey taken at the end of the semester. In this study, the SI participants earned significantly higher course grades than both groups of non-participants. However, contrary to expectations, the motivational control group did not outperform the other non-participants.

Motivation and Effort

A tacit assumption of the research on motivation and performance is that motivation leads to actual effort expended. According to Wolters (1998), motivation is the “willingness or desire to be engaged and commit effort to completing a task” (p. 224). In college work, motivation can be observed in action by the behavioral choices made by a student, the intensity of effort, and the persistence of activity (Pintrich & Schrauben, 1992). However, motivation is usually measured by surveys of subjects’ intentions instead of by observing behavior (Blumenfeld, 1992; Longo, Lent, & Brown, 1992).

The motivation for college students to exert effort in a specific way, such as attending SI sessions, can be viewed as a subset of the general motivation to succeed in college courses. The expectancy-value model of motivation (Eccles, Adler, Futterman, Goff, Kaczala, Meece, & Midgely, 1983; Weiner, 1986; Petri, 1995) describes the motivational conditions that lead students to apply efforts using various study strategies, including reading the text, reviewing notes, working problems, and attending SI sessions. Under this model, the motivation for students taking particular actions (which is the same as the likelihood of students taking those actions) is related to both the value that they place on an outcome and the expected contribution toward the outcome that they perceive the action will provide. Expectancy-value theory applied to Supplemental Instruction would hold that students’ motivations for attending the sessions are due to the high value they place on good grades and their belief that they might earn higher grades by participating in SI sessions. Students motivated to attend SI sessions and students generally motivated to succeed share the high value placed on good grades, but may differ on benefit expectancies for different study strategies.

The Setting of This Study

This report describes a motivational control study of SI-supported instruction at North Carolina State University, a large research institution in the

performance and falsely attribute these differences to the program being assessed.

One way to investigate the validity of assessments that use a volunteer sample as an experimental group is to use motivation as a control variable in addition to the usual control group of non-participants. The idea is to find a group of students who do not use the support service but who are thought to possess some of the same motivational qualities as the students who do use the service. This is a report of such a motivational control study with the popular academic support program, Supplemental Instruction (SI).

Supplemental Instruction

At many colleges, out-of-class study is supported by scheduled Supplemental Instruction (SI) sessions (Blanc, DeBuhr, & Martin, 1983). SI sessions provide opportunities for students to get together with fellow students to compare notes, discuss concepts, work problems, and develop strategies for studying the material. Fellow students previously enrolled and successful in the course and trained to follow the “SI model” lead the sessions. These “SI leaders” attend class, take notes, and do homework assignments in preparation for the sessions.

Supplemental Instruction is a specific review session methodology that has been validated as an exemplary program by the U. S. Department of Education (Blanc, DeBuhr, & Martin, 1983). Instead of focusing on developmental or struggling students, the SI program targets “high-risk” courses, which are defined as courses with relatively high failure and withdrawal rates. The program is not viewed as remedial, and all students are encouraged to attend. In the informal SI sessions, leaders help students find their own answers and encourage them to help one another. In this way, students are empowered with learning skills that help them succeed in the current course and in future courses.

Analyses of grade results have shown that students participating in SI achieve higher average course grades, even when controlling for prior academic achievement (Blanc, DeBuhr, & Martin, 1983; Kochenour, Jolley, Kaup, Patrick, Roach, & Wenzler, 1997). These studies did not address the possibility that the students who self-selected the SI treatment might be more motivated to succeed and put more effort into their studies.

Kenney (1989) addressed the question of whether the success of SI was simply due to the additional time-on-task in the sessions. In a direct comparison with the more traditional problem sessions used to support mathematics lectures, sessions operated under the SI methodology were found to result in higher achievement.