ETS TOEFL

Quality Beyond Measure.

ISSN 1930-9317

TOEFL iBT™ Research Report
TOEFL iBT–14

The Effectiveness of Feedback for L1-English and L2-Writing Development: A Meta-Analysis

Douglas Biber

Tatiana Nekrasova

Brad Horn

February 2011

The Effectiveness of Feedback for L1-English and L2-Writing Development: A Meta-Analysis

Douglas Biber, Tatiana Nekrasova, and Brad Horn
Northern Arizona University

RR-11-05

ETS is an Equal Opportunity/Affirmative Action Employer.

As part of its educational and social mission and in fulfilling the organization's non-profit Charter and Bylaws, ETS has and continues to learn from and also to lead research that furthers educational and measurement research to advance quality and equity in education and assessment for all users of the organization's products and services.

Abstract

This research project undertook a review and synthesis of previous research on the effectiveness of feedback for individual writing development. The work plan was divided into two main phases. First, we surveyed all available studies that have investigated the effectiveness of writing feedback, including both quantitative and qualitative research, for students who have learned English as a first language (L1-English), students who have learned English as a second language (L2-English), and students who have learned second languages other than English. The results of this survey are described in a narrative overview of previous research pertaining to the role of feedback in the development of writing proficiency. The survey also identified the major theoretical constructs used in this research domain, providing the basis for subsequent statistical analysis.

Second, we built on this survey to carry out a meta-analysis of empirical studies in this research area. The goal of the meta-analysis was to provide a quantitative investigation of the extent and ways in which feedback has been effective, summarizing the findings of previous quantitative studies that have employed suitable statistical measures. Several analytical steps were required for the meta-analysis: developing a coding rubric; analyzing the research design and adequacy of reporting in studies to determine if they were suitable for inclusion; coding each study for all relevant research design factors; computing effect sizes for each study; and analyzing and interpreting the general patterns that hold across this set of studies.

The meta-analysis compared the gains in writing development with respect to several different kinds of feedback. Overall, feedback was found to result in gains in writing development. Beyond that, there were several predictable findings (e.g., that written feedback is more effective than oral feedback for writing development) and several other more noteworthy trends (e.g., that peer feedback is more effective than teacher feedback for L2-English students; commenting is more effective than error location; and in general, focus on form and content seems to be more effective than an exclusive focus on form).

Key words: feedback, writing development, meta analysis, commenting, error analysis

The Test of English as a Foreign Language™ (TOEFL®) was developed in 1963 by the National Council on the Testing of English as a Foreign Language. The Council was formed through the cooperative effort of more than 30 public and private organizations concerned with testing the English proficiency of nonnative speakers of the language applying for admission to institutions in the United States. In 1965, Educational Testing Service (ETS) and the College Board® assumed joint responsibility for the program. In 1973, a cooperative arrangement for the operation of the program was entered into by ETS, the College Board, and the Graduate Record Examinations® (GRE®) Board. The membership of the College Board is composed of schools, colleges, school systems, and educational associations; GRE Board members are associated with graduate education. The test is now wholly owned and operated by ETS.

ETS administers the TOEFL program under the general direction of a policy board that was established by, and is affiliated with, the sponsoring organizations. Members of the TOEFL Board (previously the Policy Council) represent the College Board, the GRE Board, and such institutions and agencies as graduate schools of business, two-year colleges, and nonprofit educational exchange agencies.

❖ ❖ ❖

Since its inception in 1963, the TOEFL has evolved from a paper-based test to a computer-based test and, in 2005, to an Internet-based test, TOEFL iBT™. One constant throughout this evolution has been a continuing program of research related to the TOEFL test. From 1977 to 2005, nearly 100 research and technical reports on the early versions of TOEFL were published. In 1997, a monograph series that laid the groundwork for the development of TOEFL iBT was launched. With the release of TOEFL iBT, a TOEFL iBT report series has been introduced.

Currently this research is carried out in consultation with the TOEFL Committee of Examiners. Its members include representatives of the TOEFL Board and distinguished English as a second language specialists from the academic community. The Committee advises the TOEFL program about research needs and, through the research subcommittee, solicits, reviews, and approves proposals for funding and reports for publication. Members of the Committee of Examiners serve four-year terms at the invitation of the Board; the chair of the committee serves on the Board.

Current (2010-2011) members of the TOEFL Committee of Examiners are:

Alister Cumming (Chair)	University of Toronto
Carol A. Chapelle	Iowa State University
Barbara Hoekje	Drexel University
Ari Huhta	University of Jyväskylä, Finland
John M. Norris	University of Hawaii at Manoa
James Purpura	Columbia University
Carsten Roever	University of Melbourne
Steve Ross	University of Maryland
Mikyuki Sasaki	Nagoya Gakuin University
Norbert Schmitt	University of Nottingham
Robert Schoonen	University of Amsterdam
Ling Shi	University of British Columbia

To obtain more information about the TOEFL programs and services, use one of the following:

E-mail: toefl@ets.org
Web site: www.ets.org/toefl

Acknowledgments

We would like to thank several anonymous reviewers for their comments on previous drafts of this report, and especially Yasuyo Sawaki for detailed comments and suggestions for revision. We also are appreciative to Brian Feliciano and Xiaoming Xi at ETS for their continuing help through all stages of the project.

Table of Contents

Page

List of Tables

List of Figures

Page

1. Introduction

Feedback is generally regarded as essential for writing development at all levels, from students at the kindergarten through 12th grade (K–12) levels, to college freshman taking composition courses, to graduate students working on dissertation projects. Similarly, feedback has been considered essential for both first language (L1) and second language (L2) writing development.

Despite this widespread perception, much less agreement exists on the kinds of feedback that actually make a difference, or even on the kinds of gains in proficiency that can be expected from feedback. Numerous papers advocate one or another approach, and many other studies describe a writing course where a particular approach was used. Many other papers adopt a (quasi)experimental approach, measuring gains in writing proficiency that result from feedback.

Numerous factors must be considered in any study of feedback to determine which ones are actually influential. For example, feedback can be provided by the teacher, other students, or an automated system on a computer. Feedback can be written or spoken, and it can focus on content, organization, grammatical form, or usage (e.g., spelling). If written, feedback on form can comment on the existence of errors, identify the location of specific errors, or actually correct errors. And then, of course, questions must be addressed about how to measure potential improvements in writing performance resulting from feedback, for example, focusing on reduction in errors, the extent to which students incorporate revisions, or overall holistic assessments of writing quality.

Recently, Hyland and Hyland (2006) carried out a comprehensive survey of research on feedback, identifying several of the most important issues and describing numerous studies that investigate those issues (cf., DiPardo & Freedman, 1988). However, despite the large number of studies (over 200 in their survey), Hyland and Hyland concluded that there is surprisingly little consensus and most of the fundamental questions remain unanswered:

> While the research into feedback on L2 students' writing has increased dramatically in the last decade, it is clear that the questions posed at the beginning of this paper have not yet been completely answered. […] Nor are we a lot closer to understanding the long term effects of feedback on writing development. (p. 96)

In part, this lack of consensus results from the diverse research designs and methodologies used in previous studies of feedback. However, an additional limitation has been

the lack of quantitative techniques to document the state-of-the-art in this research domain. That is, previous survey articles, such as Hyland and Hyland (2006), have relied on descriptive narratives to survey previous research in this domain. However, those surveys provided no quantitative analysis of the distribution of research approaches and designs within the domain. For example, how many of these studies have been qualitative reports versus quantitative empirical studies? How many of these studies have used experimental designs versus other kinds of quantitative comparisons?

In fact, authors of state-of-the-art articles in applied linguistics usually pay little attention to the methods that they used themselves in carrying out the survey. That is, it has generally been assumed that the research for a survey article consists of finding as many publications on a topic as possible, determining the types of research and the main research issues represented by those studies, and then describing the studies that fall into each type. Such surveys rarely specify how articles were selected for inclusion in the review or provide any other evidence that the reader can use to evaluate the representativeness of the survey. Rather, the survey depends crucially on the expert knowledge of the authors. While such descriptions are a tremendous resource for future researchers beginning work in a particular domain, it is difficult to determine the extent to which the survey actually represents the research domain.

To address these concerns, recent research in applied linguistics has begun to advocate *systematic research syntheses*, applying the techniques of meta-analysis that have been developed over the past few decades for social science research. Systematic research syntheses differ from traditional literature surveys in three major ways (Norris & Ortega, 2006, pp. 6–7; see also Norris & Ortega, 2007, pp. 807–8):

1. The selection of studies to be included in the survey is a deliberate part of the research process, with explicit procedures for defining the population and identifying the research studies to be included or excluded from the survey.

2. Each research study is critically evaluated for the appropriateness of its research design and application of statistical procedures (rather than uncritically reporting study conclusions).

3. Each research study is analyzed with respect to the same set of design variables and values, applying a coding scheme developed for the entire meta-analysis.

A subset of the studies included in a systematic research survey will be suitable for a subsequent stage of analysis: a statistical meta-analysis based on comparison of effect sizes across studies. To be included in this stage of the research synthesis, a study must employ an experimental research design and be explicit and complete in its reporting standards. By comparing the magnitude of effect sizes across multiple studies in a research domain, it is possible to compare the importance of different factors based on the cumulative evidence of all empirical studies in the domain.

Two recent studies have applied the techniques of statistical meta-analysis to study writing development. These studies included some information on the effectiveness of feedback, although that was not the primary focus of either one. Truscott (2007) focuses on the quite restricted question of the extent to which error correction influences writing accuracy for L2-English students. This study concluded that overt error correction actually has a small negative influence on learners' abilities to write accurately. However, the meta-analysis was based on only six research studies, making it somewhat difficult to be confident about the generalizability of the findings.

The second study, Graham and Perin (2007) was much larger in scope but focused on writing instruction (for L1-English adolescent students) rather than the effectiveness of feedback. As a result, that study considered factors such as different instructional approaches (e.g., writing as product versus process); explicit instruction in grammar, sentence combining, writing strategies, and so on; prewriting activities; and the use of word-processing for writing. The only factor in that study that was directly relevant to the present inquiry was *peer assistance*, which was identified with a moderately large increase in writing quality.

The present report focuses exclusively on the influences of feedback for writing development, providing a large-scale and systematic synthesis of research on this topic. Because of the need to follow explicit procedures at all stages, the report is organized somewhat differently from a traditional literature review. In Section 2, we document the procedures that we used to describe the research domain and to attempt to construct an exhaustive catalog of research studies within that domain. We then describe our initial coding scheme, identifying the major ways in which studies of writing feedback can differ from one another. We also discuss the research designs and reporting standards that are required for a study to be suitable for inclusion in the statistical meta-analysis.

In Section 3, we provide an empirical survey of research studies in this domain, including discussion of the breakdown of studies across the major variables included in our coding scheme. Section 3 also describes the subset of studies that are appropriate for statistical meta-analysis of their effect sizes.

In Section 4, we turn to the procedures used for the statistical meta-analysis. Numerous analytical decisions are required for this stage of the synthesis, and our goal here is to describe those as fully and explicitly as possible.

Section 5 provides the most important information from this synthesis: the results of the statistical meta-analysis. In this section, we compare the magnitude (and dispersion) for the effect sizes of several different factors that have been hypothesized to influence the effectiveness of feedback for writing development. Based on these analyses, we are able to provide an overall perspective on the influence of feedback, identifying factors that seem to make a difference for writing development and those that seem to be less influential.

A summary and discussion of the statistical meta-analysis is taken up in Section 6.

2. Procedures I: Describing the Research Domain

The first major stage of this project was to describe the research domain. Research for this stage was carried out in three steps: First, we conducted a literature search to identify all relevant studies, employing the procedures described in Section 2.1. Second, we developed an explicit coding scheme, described in Section 2.2, that included all major variables represented in the research designs of these studies and the major values that were distinguished for those variables. Finally, we coded each research study for all variables in our coding scheme.

2.1. The Literature Search

The literature search began with an operational definition of the population of studies, followed by a comprehensive sampling of studies in that population. For the purposes of this search, we attempted to identify all studies that addressed the central research question of our research synthesis:

Which kinds of feedback are influential for which kinds of gains in writing proficiency?

This research question has two main components: (a) the different operationalizations of feedback (kinds of feedback) and (b) the range of outcome measures (kinds of gains in writing

proficiency). Thus, we included articles that investigated different sources of feedback (e.g., teacher, peer, computer), as well as different forms of feedback (e.g., direct correction, editing codes, highlighting) delivered in different modes (i.e., spoken, written, and computer mediated). For similar reasons, in addition to articles that report development in terms of writing proficiency measures, we also included articles that reported results from other outcome measures (e.g., surveys of student attitudes, analyses of post-feedback revisions).

We included studies of both native and/or nonnative English speaking students (including developing and remedial writers). Our goal in doing this was to allow comparison of the two populations, asking whether feedback is influential in the same ways and to the same extent in L1 and L2 populations.

Location and selection of research studies. The first step in our survey was to identify research journals that could potentially publish articles on feedback. This was done by exploring library catalogs and databases and by including any journal cited in previous survey studies. The following journals were included in this step:

Academic Writing Across the Disciplines
Applied Linguistics
Assessing Writing
Australian Journal of Language and Literacy
British Journal of Educational Technology
CALICO Journal
CALL Electronic Journal
Canadian Modern Language Review
College Composition and Communication
Computer Assisted Language Learning
Computers and Composition
Computers & Education
ELT Journal
English for Specific Purposes
English Journal
Foreign Language Annals
International Review of Applied Linguistics
Issues in Writing
Journal of Basic Writing
Journal of Educational Psychology
Journal of Educational Research
Journal of English for Academic Purposes
Journal of Second Language Studies
Journal of Second Language Writing
Jnl of Technical Writing & Communication
Language Learning
Language & Learning Across the Disciplines
Language Teaching
Language Teaching Research
Modern Language Journal
ReCALL
Research in the Teaching of English
RELC Journal
Rhetoric Review

Second Language Research
Spaan Fellow Working Papers
Studies in Second Language Acquisition
System
Teaching English in the Two Year College
TESL Canada Journal
TESL-EJ
TESOL Journal
TESOL Quarterly
Writing Center Journal
Written Communication

For each of these journals, we searched the online table of contents to identify all articles that had any of the following keywords: feedback, response, comment(ing), revision, peer, and writing. The range of dates searched was dictated by the archival status of individual journals but in general spanned the period 1980–2007. In addition, an online search of the ERIC database was conducted using the keywords *writing* and *feedback*. Our literature search focused primarily on studies published in academic journals. Further, as individual articles were being analyzed, the list of references in each was reviewed to identify additional articles (including studies published in edited books) that had not yet been collected. The studies included in our literature survey are mostly published research articles; we made no systematic attempt to include studies from the "fugitive" literature (e.g., unpublished papers, dissertations, conference presentations), apart from research papers identified through the ERIC database.

Using these methods, we were able to collect articles representing a variety of epistemological traditions. Unlike the methods used for some other meta-analyses, we did not adopt *a priori* exclusion criteria regarding research methodology (e.g., accepting only experimental or quasi-experimental studies). Instead, the articles included in our survey ranged from tightly controlled experimental studies to qualitative case studies. This inclusive approach allowed us to evaluate the maturity of the research domain before selecting empirical studies for the quantitative meta-analysis.

While articles were not excluded from the survey on the basis of research methodology, we did exclude studies that were not in the research domain of focus here. In particular, we excluded the following:

- studies focusing on oral (rather than written) production;
- studies in which computer-mediated chat was the target of feedback (because engaging in chat is a different communicative enterprise from the writing tasks normally considered in studies of writing development). (Note: we did include studies

that investigated chat as the means through which feedback on writing was delivered.); and

- studies focusing on the writing of special-needs student populations (e.g., deaf students).

2.2. Identifying the Parameters of Variation Among Research Designs and Coding the Study Reports

The central research question motivating this research synthesis has two main components: the different kinds of feedback and the different measures of improvement in writing proficiency. We thus began this project by carrying out preliminary research on how these two constructs have been approached in previous research.

Then, with that background, we developed an explicit coding rubric. The goals of this step were to identify all important factors that varied across feedback studies (e.g., age of the subjects, type of writing task required, type of feedback provided) and to itemize the possible values for each of those variables. This rubric was developed inductively, by reading through a wide sample of research studies to identify various ways in which their research designs could vary. The rubric was subsequently applied for an empirical description of this research domain (described in Section 3).

Operationalizations of feedback in the research literature. On initial consideration, feedback might seem to be a simple construct—providing a constructive evaluation of writing quality to the student. However, in actual practice and in the research literature, an extremely wide range of variation was found in the actual realization of feedback. These differences can be described with respect to five variables: type, focus, tone, mode, and source.

Type of feedback. In research on traditional teacher-generated feedback, the distinction between *direct* and *indirect feedback* has been one focus of studies in the areas of writing and second language acquisition (SLA) research (e.g., Ferris, 2003, 2006; Ferris & Roberts, 2001; Robb, Ross, & Shortreed, 1986). The term *direct feedback* is used to denote instances where the writing instructor makes an explicit correction to the student's text (e.g., by writing in the correct grammatical form), while *indirect feedback* denotes instances where the instructor indicates that something about the student's writing is problematic (e.g., by underlining an ungrammatical construction and/or marking the problematic section of text with a special code) but does not

provide an immediate correction. In actual practice, direct feedback is rarely used as a treatment in empirical research, while numerous types of indirect feedback have been investigated. These include identifying the location of problems, providing comments in the margins, global comments at the end of a paper, and even oral comments given to the student.

Focus of feedback. This area of research has dealt with the features of student writing (e.g., lexis, grammar, mechanics, organization, content) that the feedback provider chooses to focus on. As noted above, much feedback research has focused on error correction. Researchers on second language writing research distinguish between grammatical and word choice errors, because such "L2 errors" are thought to stigmatize L2 users. For example, Ferris (1999) divided such errors into two classes, which she labeled *treatable* and *untreatable*. Treatable errors are those that can be addressed through explicit instruction and include language features such as article usage and subject-verb agreement (i.e., rule-governed constructions). Untreatable errors are those that are less readily teachable in that they are not governed by a clear or simple set of rules. Problems with word choice are one example Ferris gave of such untreatable errors.

The predominating emphasis on error correction seems to be motivated by the perceived severity of different error types among readers of L2 texts. However, not all teacher comments address aspects of student language use that can be objectively characterized as incorrect or even problematic (e.g., positive feedback, clarification questions). Furthermore, many student writers desire guidance in these additional areas, especially as they reach more advanced levels of writing proficiency (Leki, 2006). While feedback on surface level errors may be comparably easy to provide (both for human teachers and computer programmers), an important question is whether this type of feedback leads to greater gains in student writing proficiency than more holistically focused feedback on text content, organization, or audience/purpose.

Tone of feedback. Following from the idea that not all feedback focuses on student errors, it is also the case that feedback can vary in the degree to which it praises areas of strength or criticizes areas of weakness (see, e.g., Hyland & Hyland, 2001). Concern has been expressed in the literature that overly negative feedback will adversely affect the student's motivation. At the same time, it is possible that some students may view positive feedback as less useful than critical feedback that identifies features of their writing that need to be revised. Thus, an important concern for instructors is determining the best tone for constructive criticism, given (a)

the nature/amount of feedback that needs to be provided and (b) the nature of the teacher-student interpersonal relationship.

Mode and source of feedback. Finally, feedback can be provided through any available channel, or *mode*: oral, written, or computer mediated. Although it has not been a major factor in previous research, several studies considered the influence of one mode of delivery over another. Similarly, feedback can be provided by the teacher or by other students, or even generated automatically by computer.

Operationalizations of writing development: Outcome measures of the effects of feedback. To demonstrate the effectiveness of feedback, researchers have used measures of writing proficiency (e.g., Chandler, 2003; Min, 2006), as well as survey instruments designed to elicit student perspectives (e.g., Ferris, 1995). Writing proficiency measures that have been used in feedback research included ratings obtained from classroom teachers and/or trained judges using holistic and analytic rating scales, as well as other measures of syntactic and lexical complexity. Student perspectives or changes in student attitudes have been elicited using both qualitative approaches (e.g., interviews) and quantitative instruments (e.g., surveys). A third approach to analyzing the effectiveness of feedback has been to tabulate the number of suggested revisions that were ultimately adopted by the student in subsequent drafts (e.g., Min, 2006).

The coding rubric. The first major stage of a systematic research synthesis is to undertake an empirical analysis of the research domain, documenting the ways in which the central research questions have been approached within that domain. This description is also required to evaluate whether the research domain is sufficiently mature to permit a statistical meta-analysis. In the present case, our preliminary reading indicated that much of the research on writing for the past two decades has eschewed quantitative methods in favor of more qualitative approaches, especially in the area of first language composition research. While qualitative work adds to our collective understanding of how students develop their writing skills, such studies cannot be included in a quantitative meta-analysis. Thus, the ultimate goal of the analysis in this stage is to determine whether enough experimental studies—with clearly documented research designs and statistical results—exist to permit the application of meta-analytic techniques.

To accomplish the empirical analysis of the research domain, it is first necessary to develop a coding rubric that itemizes all important factors that vary across feedback studies, as well as all possible values for each of those variables. This rubric is developed inductively, by

reading through a wide sample of research studies to identify various ways in which their research designs could vary.

The coding rubric developed for the present project includes 16 variables:

- Research paradigm
- Statistical analysis
- Design variables
- Target language
- Proficiency level (for L2 studies only)
- Number of student participants
- Age/grade level of student participants
- Genre of writing task(s)
- Length of writing task(s)
- Source of feedback
- Mode of feedback
- Focus of feedback
- Tone of feedback
- Type of feedback
- Outcome measures
- Specific focus for outcome measures of writing proficiency.

These variables were used to categorize the types of research studies in this research domain, and thus it was necessary to develop an exhaustive list of values for each variable. These values and variables are shown in Table 1 (with the codes used for the meta-analysis given in square brackets; see Appendices B and C).

Table 1

The Coding Rubric: Variables and Values for Each Variable

Variables	Values	Further details
Research paradigm	Quantitative	Experimental; quasi-experimental; correlational; survey
	Qualitative	Ethnographic; case study; interviews
	Mixed methods	Combination of quantitative measures and qualitative description
	Thought piece	Theoretical argument; pedagogical primer; no original empirical data
Statistical analysis	Statistical tests reported	Record statistic(s) used, including descriptive statistics
	Statistical tests not reported	
Design type	Intact group(s)	
	Random group assignment	
	One group	
	Treatment/control [TC]	
	Pretest/posttest [PP]	
	Posttest only	
	Descriptive/ex-post facto	
Target language	L1 English [E1]	English composition studies where no mention of nonnative speaker (NNS) participants is made
	L2 English [E2]	Most North American L2 writing studies
	Mixed L1 & L2 [MX]	Comprises native speaker (NS) & NNS students
	L1 Other [O1]	Students whose native language is not English, learning to write in their native language (e.g., a study of the composing processes of Dutch L1 children)

Variables	Values	Further details
	L2 Other [O2]	Students whose native language is English, learning to write in a second language (e.g., a study of U.S. college students learning L2 Spanish composition)
Proficiency level (for L2 studies only)	Proficiency level reported	[L] = Low; [H] = High
	Proficiency level not reported	
Number of student participants	*N*-size reported	Number of participants reported in study
	N-size not reported	
Age/grade level of student participants	Age or grade level reported	
	Age not reported	
Genre of writing Task	Correspondence	Business letters; personal letters; email; memos; faxes
	Creative	Fiction; poetry
	Pedagogical [PD]	"Learning" genres, such as five-paragraph essays
	Personal	Diaries; journals; reflective essays
	Research/academic	Scientific research articles; dissertations; theses; term papers
	Other genres [O]	Legal writing; journalism
	Genre not reported	
Length of writing Task	Length reported	
	Length not reported	
Source of feedback	Teacher [TE]	Feedback from course instructor
	Peer [PE]	Feedback from another student
	Tutor	Feedback from writing center tutor

Variables	Values	Further details
	Student	Self-correction
	Computer	Computer-generated feedback (not just computer-*mediated* feedback)
	Other [O]	
	Source not reported	
Mode of feedback	Oral [OR]	Face-to-face conferencing; tape-recorded comments
	Written [WR]	Marginal comments; end comments; editing codes; circles/underlines
	Computer-mediated [CM]	Internet chat; email
	Mode not reported	
Focus of feedback	Grammar	Subject-verb agreement errors; tense/aspect errors; pedagogical grammar issues in L1 studies
	Vocabulary	Collocation errors; other word choice issues
	Spelling	Spelling errors
	Organization [O]	Topic sentence; discourse markers; transitions; paragraphing; conclusion; *order* of content
	Content [C]	*Correctness* of content; *completeness* of content
	Punctuation / mechanics	Comma errors; end punctuation errors; indentation; capitalization; but *not* spelling
	Other	Anything not captured by the other values for this variable
	Form [F]	Grammar, spelling, punctuation
	Content and form [C,F]	Content and form
	Focus not reported / specified	
Tone of feedback	Negative	Comments on what the student has done wrong

Variables	Values	Further details
	Positive	Comments on what the student has done right
	Mixed	Comments on both strengths and weaknesses of text
	Tone not reported/specified	
Type of feedback	Location of error/problem/issue indicated [LO]	Location of an error is marked (circled, underlined), but no feedback is given on why it is an error or how it might be corrected
	Comment [CM]	Teacher/peer writes prose comments in the margin or at the end of the paper
	Other [O]	Other types of feedback, including direct correction/reformulation [DC]; editing codes [EC], error existence [EX], metalingusitic explanation of an error [ML], spoken explicit comments [SE], spoken implicit comments [SI]
	Multiple [M]	Multiple types of feedback are provided, such as both location and explanatory comments
Outcome measures	Writing proficiency measures [WP]	Holistic ratings of writing quality, measures of spelling accuracy, grammatical accuracy
	Attitude measures	Likert-scale items
	Records of composition strategies/processes employed	Records of time spent planning, drafting, etc.; eye-tracking records
	Records of revisions [RV]	Number/extent of revisions made
	Other [O]	
Focus for outcome measures of writing proficiency	Grammar [GR]	
	Spelling [SP]	
	Holistic [H]	
	Content [C]	

Most studies included in our study involved revisions made to an essay based on the same prompt over a period of time in response to different kinds of feedback. (McGroarty and Zhu 1997 was exceptional in this regard, because they evaluated writing development across essays based on different prompts.) The outcome measure for most quantitative studies was a measure of writing proficiency (either holistic quality or grammatical accuracy) based on evaluation of the final (revised) written product. However, in a few cases, studies simply documented the extent to which a student made any revisions, regardless of the contributions those revisions made to the quality of the final product.

Coding the studies. The initial coding of studies for general variables, such as the research approach, general design type, and target population, was carried out by the second and third authors of the report (TN and BH). Any controversial coding decisions were discussed by all three authors and resolved through consensus.

Subsequently, more detailed coding was undertaken by the second author (TN) for the purposes of the quantitative meta-analysis. The first step for this process was to identify the sub-set of studies that were suitable for inclusion in the analysis: studies that were published in the last 25 years, used quantitative measures, had an experimental (or at least quasi-experimental) design, were explicit about the types of feedback that were provided, employed a clear basis for comparison, and included an outcome indicator that measured change in students' writing proficiency or behavior (see Section 4.1 below). Any controversial coding decisions during this process were resolved through consensus by discussion between the first two authors (DB and TN).

3. Empirical Survey of the Research Domain

Based on the sampling methods described in Section 2.1, we collected a total of 306 articles that addressed the effectiveness of feedback for writing development. Our goal here was to obtain an exhaustive sample of studies published in the last 30 years, resulting in a much larger collection of publications than in some previous meta-analyses.

Figure 1 shows the breakdown of studies across year of publication. The trend here shows a dramatic increase in the number of feedback studies over the past 30 years. This trend reflects two factors. First is the general information explosion, with an increase in the number of academic journals and publications in all disciplines, and the more specific increase in the number of studies investigating the effect of feedback. Second, and more important for us here, is that this increase suggests researchers (and teachers) have shifted away from an uncritical

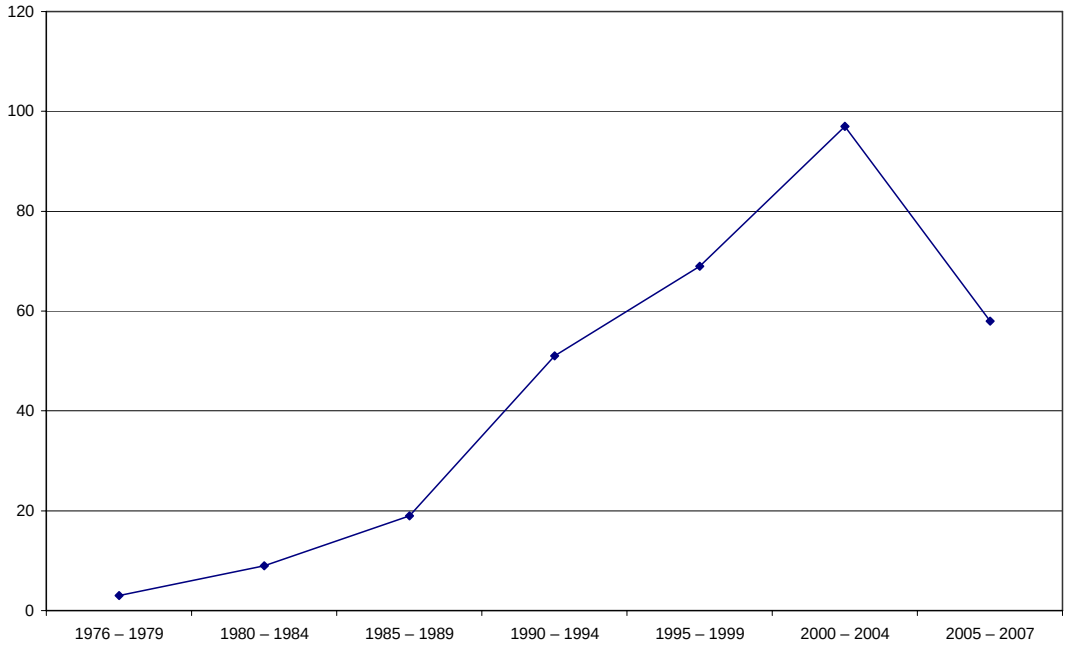

Figure 1. **Number of publications by date.**

belief in the effectiveness of feedback toward a recognition that feedback can take many different shapes and its effectiveness needs to be studied in its own right. (The last period includes only 2.5 years, accounting for the apparent decrease in publications.)

Figure 2 shows that studies in this domain have adopted the full range of research methodologies, with both qualitative and quantitative approaches represented by a large number of studies. In addition, numerous *thought pieces*—either survey articles describing previous research on feedback or general discussion articles—are included.

Figure 3 shows that the relative preference for one or another research approach has remained relatively constant across time, with quantitative studies being slightly more common than qualitative studies. The one notable departure from this pattern is in the period 2000–2004, which showed a dramatic increase in the number of qualitative studies while the number of quantitative studies remained constant. This shift might reflect a more general paradigm shift influenced by postmodern thinking in general, valuing ethnographic reports of individual case studies over reports of the general trends in a large sample of individuals. Because comparatively few studies are included in the most recent period, it is not clear whether this trend continues.

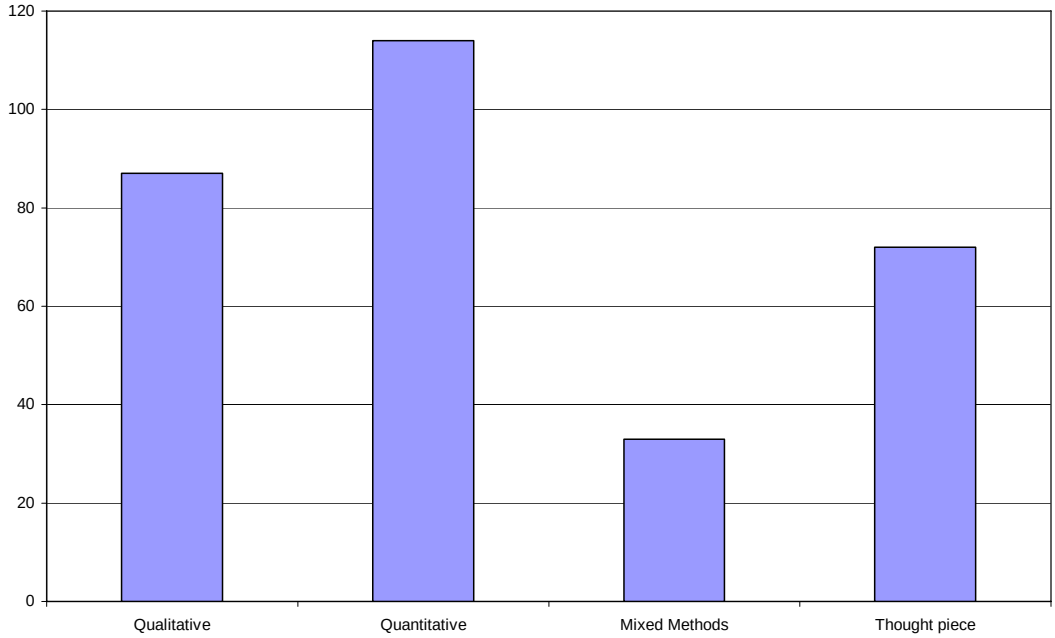

Figure 2. **Number of research publications by research approach.**

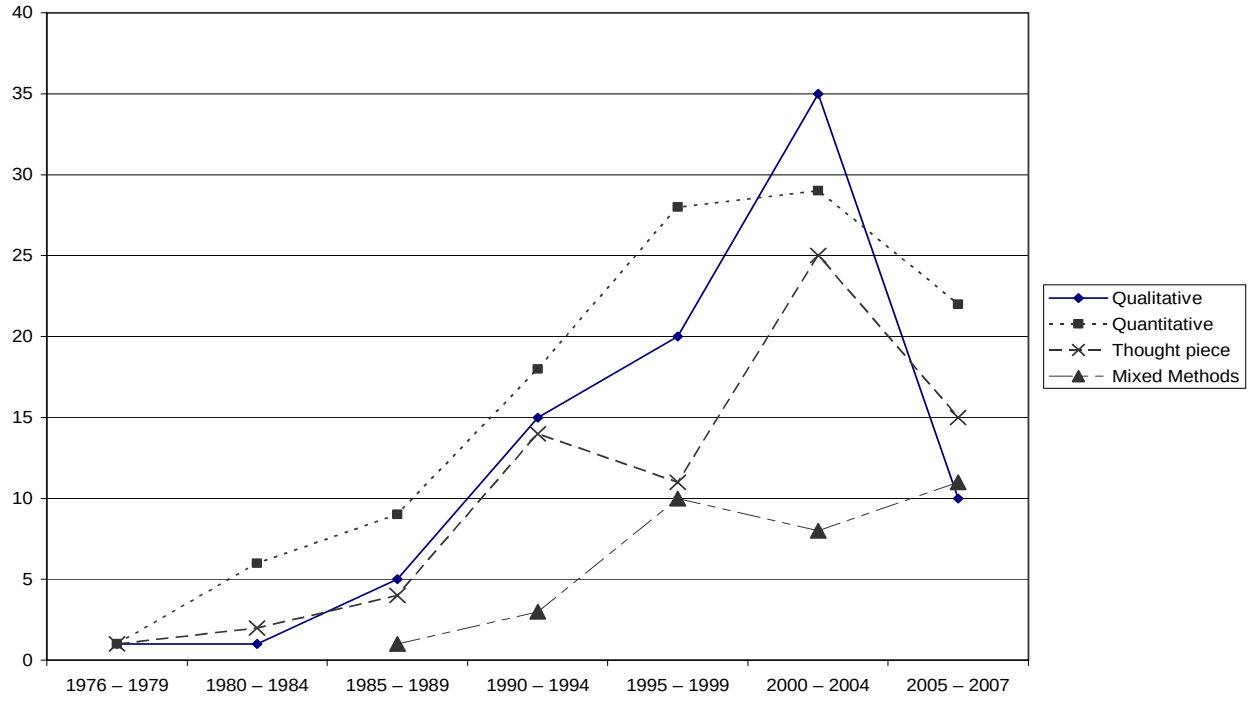

Figure 3. **Number of publications from each research approach by date.**

Studies have also varied in the target population that has been investigated (although many studies do not provide full details on the subjects). For example, Figure 4 shows that learners of English have been the primary target of investigation, although there have also been numerous studies of feedback that focused on the writing development of native English speakers. However, there has been a shift in research focus across time, as shown in Figure 5: Through the 1980s, equal interest was found in the influence of feedback for both L1 and L2 learners of English (although the number of studies is comparatively few). However, by the mid-1990s, a dramatic shift in focus occurs with many more studies focusing on learners of English than on the writing development of native English speakers.

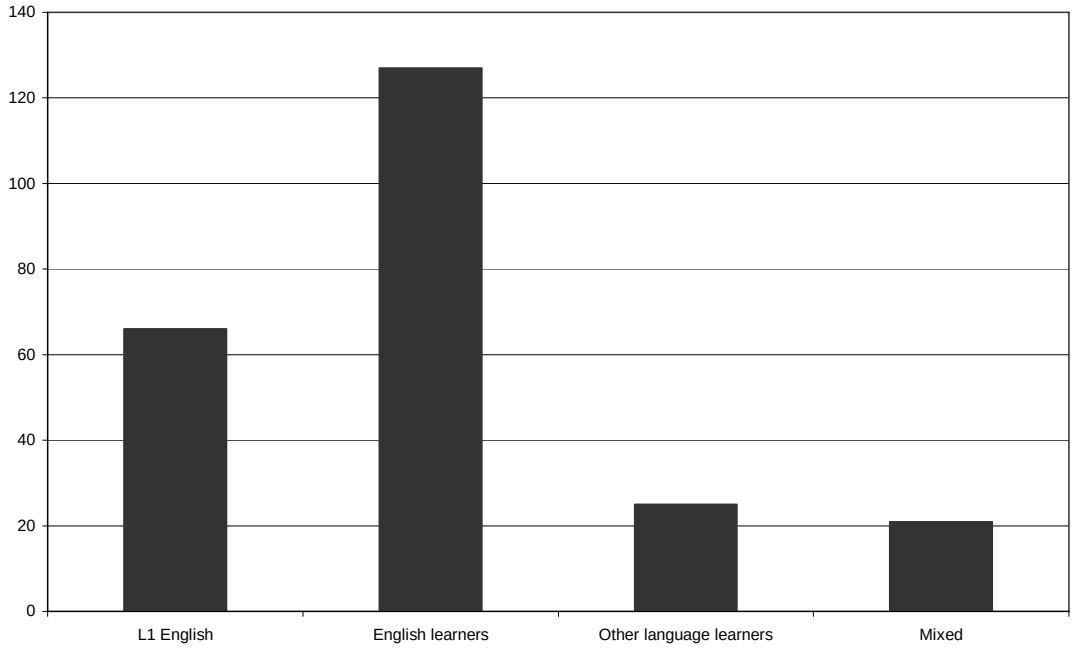

Figure 4. **Number of publications by target population**

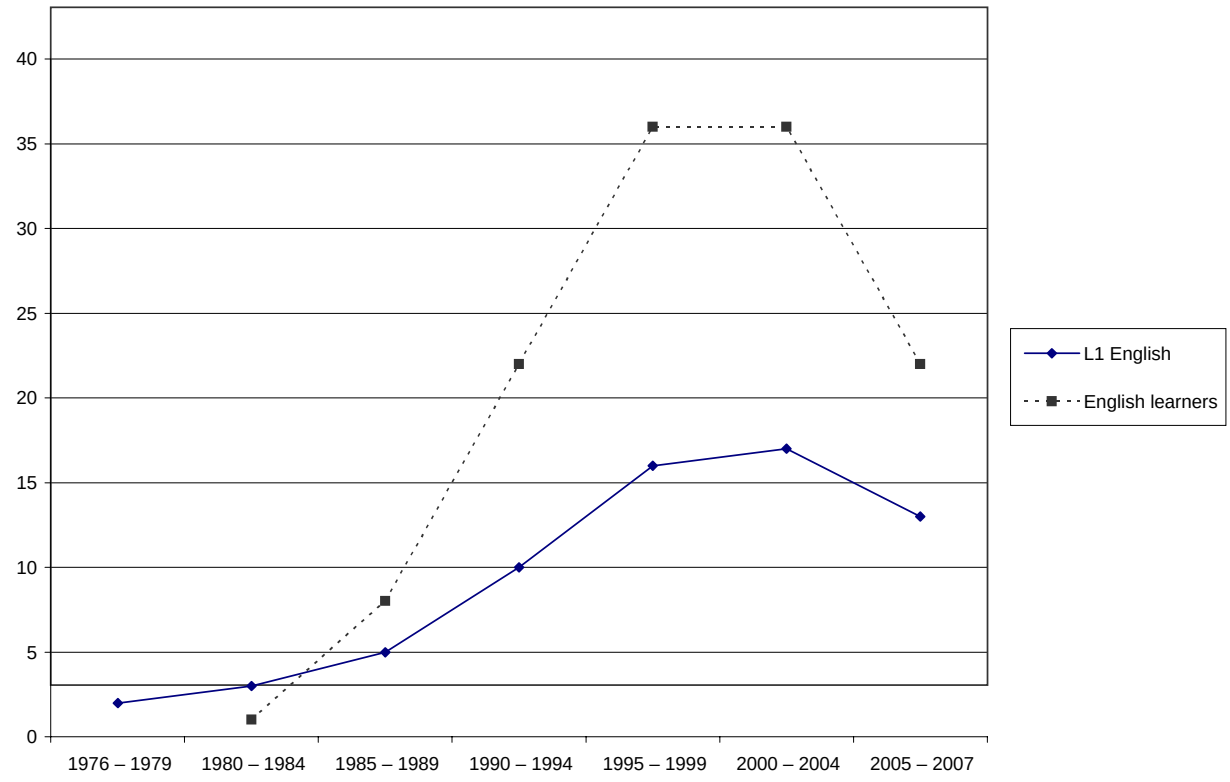

Figure 5. **Number of publications focusing on L1 versus L2 learners by date.**

Feedback studies that were focused on learners of English investigated the full range of proficiency levels, as shown in Table 2.

Table 2

Breakdown of Studies by Proficiency Level of the Target Population (Includes Only Studies of English Learners)

Proficiency levels	Number of studies
Low/beginner	17
Intermediate	19
Advanced	26
Mixed	21
Unspecified	44
Total	122

Table 3 shows that the majority of feedback studies (whether L1 or L2) have focused on the writing of college-aged students, while comparatively few studies have investigated the influence of feedback for younger students.

Turning to the nature of the student writing, Table 4 shows that the overwhelming majority of feedback studies have used pedagogical writing tasks, such as the five-paragraph essay.

Table 3

Breakdown of Studies by Age of the Target Population

Target population age	Number of studies
Ages 4–9	8
Ages 10–12	10
Ages 13–18	15
College-age	159
Other adult ages	29
Unspecified	85
Total	306

Table 4

Breakdown of Studies by the Genre of the Writing Task

Genre of writing task	Number of studies
Pedagogical	187
Personal correspondence	7
Personal journal or diary	7
Creative writing	2
Research/academic writing	8
Other/unspecified	95
Total	306

Since the central research question for this project focuses on the effectiveness of feedback, we coded five different variables to capture the different ways in which feedback was realized: source, mode, focus, type, and tone. As Table 5 shows, the large majority of these studies focus on feedback given by the teacher. In this regard, feedback studies probably reflected typical classroom practice, but they were at odds with many theoretical discussions that advocated the utility of peer feedback.

Most feedback on student writing was communicated in writing, either using a computer (32 studies) or with feedback written by hand (94 studies). Many studies did not report the mode of feedback; only 37 studies reported giving oral feedback, and an additional 36 studies used multiple modes.

Most of the studies that reported on the focus of feedback compared the influence of multiple categories (80 studies—usually a focus on both form and content). However, most studies did not report a specific focus, while only 14 studies had a single focus: 3 on content, 9 on grammatical form, and 2 on spelling.

Similar to the incomplete reporting typical of the other parameters, most studies in our sample did not report the particular type of feedback. For the remaining studies, Table 6 shows that written comments are the most common type of feedback, while a large number are also based on multiple types of feedback.

Table 5

Breakdown of Studies by the Source of Feedback

Source of feedback	Number of studies
Computer	18
Peer/other students	34
Self criticism	17
Peers + self	6
Teacher	119
Teacher + other	38
Other/unspecified	74
Total	306

Table 6

Breakdown of Studies by the Type of Feedback

Type of feedback	Number of studies
Comments	66
Error code	2
Direct correction of error	11
Location of error	4
Multiple types	35
Other/unspecified	188
Total	306

Only 17 of the 306 studies in our sample reported on the tone of feedback. Of those, 15 claimed to provide both positive and negative feedback, and 2 provided only positive feedback. It seems unlikely that this emphasis on positive feedback was equally typical for the 289 studies that did not report on tone.

Finally, we noted in Section 2.1 above that the central research question motivating this research synthesis has two main components: the kinds of feedback (described in the preceding paragraphs) and the resulting gains in writing proficiency. Table 7 shows that there is very little agreement on the best way to operationalize writing proficiency or development. 102 of the studies in our sample provided no specific measure of writing development. The remaining 204 studies, though, used a wide range of different measures, including questionnaires to determine student attitudes, direct comments on progress from teachers or students, and a record of the extent to which essays have been revised. Surprisingly few studies included a direct measure of writing quality, which might include scores for grammatical accuracy, content, organization, or an overall holistic rating for quality.

In sum, our survey of research relating to feedback on writing shows that considerable depth exists in this research domain, with numerous studies undertaken from multiple perspectives. About half the studies in this research domain have been quantitative, and those studies have included many variants of research design. There are advantages to this diversity, in

Table 7

Breakdown of Studies by the Outcome Measures of Writing Development

Outcomes used to measure writing development	Number of studies
Attitude measures	35
Revisions	39
Attitude measures plus revisions	12
Composition strategies	7
Composition strategies plus revisions	3
Essay score for quality or grammatical accuracy	40
Essay score plus attitude measures	18
Essay score plus revisions	18
Teacher or student comments on progress	32
Other/unspecified	102
Total	306

that each new research study considers slightly different research questions from preceding studies. For the purposes of a quantitative meta-analysis, however, this diversity, which depends on the existence of multiple studies that are directly comparable, also presents disadvantages.

The following section turns to the methods of meta-analysis and an evaluation of this research domain to determine if it is suitable for this approach.

4. Procedures II: The Quantitative Meta-Analysis

The meta-analysis proceeded in three major steps:

1. All publications in the larger sample were examined to identify the set of studies that were suitable for inclusion in a meta-analysis.
2. Effect sizes were computed for the outcome variables in each of those studies.
3. Mean effect sizes were computed for each treatment variable as the basis for determining the influence of different forms of feedback on writing development.

We describe each of these methodological steps in turn in the following subsections.

4.1. Identifying the Subset of Studies That Are Suitable for Meta-Analysis

During the coding of research articles described in Sections 2 and 3, we made an initial determination of whether a study was potentially suitable for inclusion in the quantitative meta-analysis. There were four major requirements for this initial screening (following the procedures used in Norris & Ortega, 2000, pp. 432–33):

1. The study was published in the last 25 years (between 1982 and 2007).
2. The study used quantitative measures and had an experimental (or at least quasi-experimental) design. Specifically, the study had to use and report on quantitative measures of effectiveness, for specific types of feedback.
3. The independent variables measured feedback characteristics, including source of feedback (e.g., teacher, peer, tutor, student, computer), focus of feedback (e.g., grammar, vocabulary, spelling, organization, content, mechanics, rhetorical organization), or type of feedback (direct comment, editing code, rating, etc.).
4. The dependent variables included an outcome indicator that measured the impact of specific types of feedback on participants' writing behavior, including writing proficiency (e.g., grammatical accuracy or holistic quality rating), increase in text length, attitude, strategies/processes employed, number/extent of revisions made.

Based on these criteria, 112 studies were identified as potentially relevant for the meta-analysis. These studies were then subjected to a second round of closer scrutiny to determine if the design and reporting standards were in fact adequate for our purposes here. Unfortunately, it turned out that a large number of additional studies were excluded in this second phase, for the following reasons:

1. The research design was not suitable for inclusion. That is, following Norris and Ortega (2000), we included only studies with designs based on mean differences: a pretest/posttest design, or a control group/experimental group design.

   Several studies were excluded because they used correlational designs. Although it is possible to compute effect sizes from such designs, these studies addressed fundamentally different kinds of research questions, and so they could not be readily compared to the effect sizes from group-comparison studies. Twenty-four studies were excluded for this reason.

2. The study addressed a different research question from the one that we are investigating in our project (e.g., studies on whether males/females produced more errors or studies on whether grading rubrics are biased to favor males or females). Some of these studies had a pre-post test design, but no actual feedback was provided. Sixteen studies were excluded for this reason.

3. The study was incomplete in its reporting of the design, sample, or statistical findings. Specifically, to be included in the meta-analysis, the study must report one of the following: (a) the sample size, mean scores, and standard deviations for each group, (b) between-groups *t* or *F* values together with *df*, or (c) individual scores on outcome measures for all participants. Twenty-four studies failed to meet these reporting standards for statistical tests (e.g., reporting only significance with no *df* or no *t* value, or reporting mean scores with no standard deviation); these studies were thus excluded.

4. The study provided no clear basis for comparison. These were mostly studies of a single group that reported only posttest results. Fourteen studies were excluded for this reason.

5. The study compared multiple treatment groups with respect to a single posttest with no pretest and no control group. For example, one group received feedback on content, while a second group received feedback on form; or one group received direct correction of errors, while only general comments were provided to a second group. Although these studies addressed some of the central research issues of our synthesis, they could not be included in the meta-analysis because it was not possible to isolate the influence of individual factors. As Norris and Ortega (2006) noted: "Direct comparisons between treatment conditions are not made, because they would be idiosyncratic to the particular study, and therefore not comparable with other studies that did not operationalize the same two treatments" (pp. 27–28). Eleven studies were excluded for this reason.

In sum, 89 additional research studies were excluded at this stage, leaving only 23 published papers (reporting on 25 different studies) that were directly comparable and otherwise suitable for

inclusion in the meta-analysis. At this point, the large majority of studies in this research domain were noted as not suitable for inclusion in a meta-analysis for three general reasons:

1. Many studies in this domain were qualitative (and often anecdotal), or thought pieces, based on researchers' observations and perceptions.
2. Many of the quantitative studies were not carefully designed, or the reporting standards were not adequate for the purposes of meta-analysis.
3. Several studies were carefully designed and implemented, but they simply addressed different research questions from the one this study focuses on.

Thus, although we were able to identify a large number of research studies in our initial survey (306 studies), relatively few of these could be used in the subsequent meta-analysis (only 25 studies).[1]

4.2. Computing Effect Sizes for the Outcome Variables

The second step in the meta-analysis was to compute an effect size for each outcome variable that reflects the influence of feedback. Again following Norris and Ortega (2000, 2006), Cohen's *d*-index was selected as the most appropriate effect size estimate and calculated for each finding related to feedback that was reported with sufficient data. Cohen's *d* represents the size or importance of a difference, either between a treatment group and a control group, or between a pretest and a posttest. (Correlational designs were not included in the final meta-analysis because they are not comparable to group comparisons designs.) In either case, this difference is interpreted as reflecting the influence of some treatment. Cohen's *d* is essentially a kind of standard score representing standard deviation units. It is calculated for a specific outcome measure by subtracting the mean scores for the two groups and then dividing this difference by the pooled standard deviation of the two groups. (There are numerous reference works that provide specific formulae to be used for the computation of effect size from different primary statistics; see e.g., Cohen, 1988; Lipsey & Wilson, 2001; Norris & Ortega, 2000; Rosenthal, Rosnow, & Rubin, 2000).

The studies included in our meta-analysis are about evenly split between studies with treatment-control designs and studies based on comparison of a pretest and posttest given to a single group (i.e., with no control group; see Section 5.1 below). Treatment-control designs (independent samples) are much stronger, allowing the researcher to isolate the influence of

feedback (the treatment) apart from other factors. In contrast, pretest versus posttest (dependent samples) designs that include only a single group are relatively weak because there is no control for the influence of natural development that occurs over the course of the study (see Section 5.2 below). For this reason, our analyses in the following sections distinguish between these two design types to the extent that it was feasible, reporting separate mean effect sizes for each type. In general, the results are consistent across both treatment-control and pretest-posttest designs, but the results for the latter should be interpreted with caution.

The computation of effect size also differs for the two design types (although both are referred to as Cohen's *d*). For studies that employed a treatment-control design, we used an online calculator to compute effect sizes (Becker, 1999):

`http://web.uccs.edu/`lbecker`/Psy590/escalc3.htm`

This calculator uses the following standard formula (which is consistent with Cohen's (1988, p. 44) formula):

$$\text{Cohen's } d = (M_1 - M_2)/ \sigma_{pooled}$$

$$\text{where } \sigma_{pooled} = \sqrt{[(\sigma_1^2 + \sigma_2^2) / 2]}$$

The reporting standards for all treatment-control studies included in our sample were high, so all studies included the descriptive statistics required for the formula (i.e., the mean scores and standard deviations for each group).

Studies that employ dependent-sample designs (i.e., pretest/posttest designs) are more problematic. In theory, the original individual scores for each subject should be used to compute effect sizes for this type of design. That is, the appropriate formula for computing the pooled standard deviation for dependent samples designs is given below (Dunlap, Cortina, Vaslow, & Burke, 1996):

$$\sigma = \sqrt{[\sum(X - M)^2 / N]}$$

An alternative approach, advocated by Lipsey and Wilson (2001, p. 41–51), is to use the mean scores and standard deviations for Time 1 and Time 2 to calculate the pooled standard deviations and effect sizes.

However, none of the studies included in the present meta-analysis provided original subjects' scores, and most of the dependent-sample studies neglected to report the standard deviations for Time 1 and Time 2. This situation arises frequently in meta-analyses, and one practical solution has been to use a simplified independent samples formula—the *t* score divided by the square root of *N*—as an estimate of effect size for the dependent-sample designs. This approach, which has been followed in studies like Norris and Ortega's (2000) meta-analysis on the effectiveness of L2 instruction, was adopted in the present study. However, it has been shown that computing Cohen's *d* from the *t* score and sample size results is an overestimate of the true magnitude of the effect size (see Dunlap et al., 1996), providing an additional reason why the results for studies with dependent-sample designs should be interpreted with caution in our analysis. We therefore report the results for dependent-sample designs separately from the results for true experimental designs.

In practical terms, a Cohen's *d* of 1.0 means that the treatment group scored one standard deviation higher than the control group (or that there was a gain of one standard deviation from the pretest to the posttest.) Thus, converting all statistical differences to standard deviation units makes it possible to directly compare outcomes across studies.

No absolute standards are used to interpret effect sizes. The most widely accepted rule of thumb, proposed by Cohen (1988), is based on a survey of the typical findings in social science research: Effect sizes of $d < 0.20$ are interpreted as *insignificant*; values of *d* between 0.20 and 0.50 are interpreted as *small effects*; values of *d* between 0.50 and 0.80 are interpreted as *medium effects*; and values of *d* larger than 0.80 are interpreted as *large effects*. However, Norris and Ortega (2006, pp. 33–34) advocated a stricter standard, based in part on their findings in Norris and Ortega (2000), where it seemed that effect sizes around 1.0 were more typical for L2 instructional treatment studies.

Dunlap et al. (1996) showed that effect size estimates based on *correlated designs* (i.e., pretest versus posttest designs) will systematically overestimate the true effect, unless adjustments are applied. Specifically, they found an overestimate by a factor of 2 for studies with a correlation of .75 for the test-retest reliabilities (the typical case; see Dunlap et al. 1996, p. 171). Using this adjusted rule of thumb results in higher required effect sizes for dependent-sample designs: $d < 0.40$ is interpreted as insignificant; *d* between 0.40 and 1.00 is interpreted as

small effects; *d* between 1.00 and 1.60 is interpreted as medium effects; and values of *d* larger than 1.60 are interpreted as large effects.

One major methodological issue for meta-analysis concerns whether it is appropriate to compute multiple effect sizes from a single publication or study. For example, many studies include multiple treatment groups (e.g., that receive different kinds of feedback) where each treatment group is compared to the same control group that received no feedback. Other studies use a single treatment group and a control group, but these two groups are compared with respect to multiple outcome measures. In cases like these, it is statistically possible to compute multiple effect sizes, one for each statistical comparison. But in that case, the effect sizes are not truly independent. Including multiple effect sizes from a study provides greater weight to that particular study, which could become a problem if that study was biased in some way.

At the same time, choosing only a single comparison from a given study fails to represent the overall findings of the study and does not provide the basis for comparisons across different meta-analyses. Thus, we decided to provide a comprehensive coverage of all comparisons reported in these studies, at the risk of including multiple comparisons based on a single group.

Specifically, we used the following approach: First, we computed an effect size for every relevant mean difference reported in these studies. In total, we computed 172 effect sizes from the 25 studies included in our meta-analysis, or on average about seven effect sizes per study. These individual effect sizes are given in Appendix B. We then analyzed the independent variables associated with each effect size, to determine whether they represented distinctions that were relevant for the purpose of our meta-analysis. In cases where two effect sizes were associated with a single configuration of independent variables, we computed an average effect size. Appendix C shows the effect sizes used for our final meta-analysis.

For example, Ashwell (2000) used a pretest-posttest design for three different groups of students. Each group received feedback focused on form and content. The different kinds of feedback were provided in different orders, but those distinctions were not relevant for the purposes of our meta-analysis. Each of the three groups was then evaluated for two outcome measures: one for grammatical accuracy and one for content. Because the distinction between grammar versus content outcomes is relevant for the purposes of our meta-analysis, these individual effect sizes were retained in the final analysis. That is, each group in the Ashwell

study was used for two different effect sizes in the final analysis: one for a grammar outcome measure and one for a content outcome measure.

The analysis of the study by Berg (1999) is relatively uncontroversial: Three different groups received feedback, contrasted with a control group that received no feedback. The groups were compared for a single outcome measure. Because the three groups were independent samples, we retained all three effect sizes in the final analysis.

In contrast, Bitchener, Young, and Cameron (2005) was based on two treatment groups, each compared with a control group for numerous outcome measures. In this case, the outcome measures (e.g., preposition use, tense use, article use) were all specific indicators of the same underlying outcome type: grammatical accuracy. In addition, each group was measured at different points in time. That is, all individual effect sizes for a group are instances of the same configuration of independent variables. As a result, 12 different effect sizes were averaged for each group, so that only two average effect sizes from this study were used in the final meta-analysis.

Appendix C shows the result of this step of the analysis, displaying each of the final effect sizes (or average effect sizes) used in our final meta-analysis. A large number of the original comparisons from these studies were specific measures of the same underlying parameter, and they were averaged for the meta-analysis. Thus, the 172 individual effect sizes that we computed were reduced to 88 effect sizes used in the final meta-analysis. However, those 88 effect sizes take into account every statistical comparison reported in the original studies.

4.3. Computing Mean Effect Sizes and Dispersion Measures

Once the effect sizes were computed for each individual study, it was possible to compute mean effect sizes for the different feedback conditions. For example, it was possible to compare the mean improvement in writing quality (the mean effect size) for students who received error correction compared to students who received global comments. This comparison was accomplished by simply computing the arithmetic mean for all effect sizes of a given type.

However, a simple comparison of mean effect sizes is not in itself very meaningful without also considering the dispersion of effect sizes around that mean. This calculation is required to determine the extent to which effect sizes vary across comparisons of a given type. For this purpose, we computed 95% confidence intervals:

CI = d ± [(95% t = distribution at k-1 df) (sd / square root(k))]

where d is the average effect size, sd is the standard deviation of all effect sizes, and k is the number of effect sizes (see Norris & Ortega, 2000, p. 187; Woods, Fletcher, & Hughes, 1986). The confidence interval provides an estimate of how well we can estimate the mean score, given the number of effect sizes used for the estimate and the range of variation among those effect sizes. The confidence interval is weighted by the t score, reflecting the fact that estimates based on relatively few studies are less precise than estimates based on a large number of studies. Smaller confidence intervals indicate that the observed effects are more robust. In particular, if the confidence interval does not include 0.0, it indicates that the mean effect size is significantly different from the null hypothesis of no effect. However, if the confidence interval includes 0.0, then no interpretation of a significant difference is possible.

Using these techniques, we computed mean effect sizes[2] and confidence intervals for the theoretically relevant comparisons within this research domain. Many of these comparisons are based on small samples and so we interpreted them with caution. For comparisons based on fewer than five effect sizes, we simply reported the individual effect sizes rather than the descriptive statistics.

5. Results of the Meta-Analysis

5.1. Breakdown of Comparisons Across Study Parameters

The studies used as the basis of the meta-analysis yielded multiple instances of comparisons for several of the key parameters of variation within this research domain, permitting further exploration of those variables. The effect sizes used for the meta-analysis were evenly split between pretest/posttest comparisons (44 effect sizes) and treatment/control group comparisons (44 effect sizes). In addition, Table 8 shows that the major parameters of variation among studies are sufficiently represented for further analysis. (The specific coding information for each of these effect sizes is given in Appendix C.) Thus, for example, multiple comparisons are presented for both L1-English students and for students studying a second or foreign language (English or other languages), including both high and low levels within the L2 group. Most of these studies targeted university students (74%) performing pedagogical writing tasks (86%). Most studies focused on the effect of teacher feedback (74%), but a moderate number focused on feedback from other sources (26%). Similarly, the majority of studies focused on

written feedback, but a moderate number of studies considered oral feedback (28%). Over half of the studies provided feedback on both form and content (58%), but a substantial number of studies provided feedback only on form (27%). Most studies provided feedback in the form of comments, often together with locating specific errors. Almost all of the studies included in the final meta-analysis used outcome measures of writing proficiency (as opposed to the number of revisions made or changes in attitudes). Beyond that, studies focused on three major areas of improvement: grammatical form (usually number of errors), content, or an overall holistic rating.

Table 8

Breakdown of the Specific Comparisons Used to Compute Outcome Effect Sizes

Independent variable	Number of effect sizes
Language background	Pretest versus posttest designs
L1-English	8
L2-English	23
L2-other than English	7
Mixed	6
Subtotal	44
	Treatment versus control designs
L1-English	12
L2-English	24
L2-other than English	7
Mixed	1
Subtotal	44
L2 proficiency level	Pretest versus posttest designs
Low/beginner	16
High/advanced	17
Subtotal	33
	Treatment versus control designs
Low/beginner	12
High/advanced	16
Subtotal	28

Independent variable	Number of effect sizes
Feedback source	Pretest versus posttest designs
Teacher	40
Other (peers or computer)	4
Subtotal	44
	Treatment versus control designs
Teacher	25
Other (peers or computer)	19
Subtotal	44
Feedback mode	Pretest versus posttest designs
Written	30
Oral	12
Missing	2
Subtotal	44
	Treatment versus control designs
Written	30
Oral	13
Missing	1
Subtotal	44
Feedback focus	Pretest versus posttest designs
Form	16
Content	2
Form and content	24
Awareness of revision process	0
Missing	2
Subtotal	44

Independent variable	Number of effect sizes
	Treatment versus control designs
Form	8
Content	2
Form and content	27
Awareness of revision process	3
Missing	4
Subtotal	44
Feedback type	Pretest versus posttest designs
Comments	16
Error location	10
Comments + location	16
Other	2
Subtotal	44
	Treatment versus control designs
Comments	26
Error location	4
Comments + location	8
Other	6
Subtotal	44
Outcome measure of writing development	
	Pretest versus posttest designs
Proficiency measure	44
Revisions	0
Other	0
Subtotal	44
	Treatment versus control designs
Proficiency measure	40
Revisions	3
Other	1
Subtotal	44

Independent variable	Number of effect sizes
Specific focus of outcome measure	Pretest versus posttest designs
Grammar/form	18
Content	8
Holistic rating of quality	18
Spelling	0
Subtotal	44
	Treatment versus control designs
Grammar/form	18
Content	11
Holistic rating of quality	9
Spelling	5
Not reported	1
Subtotal	44

5.2 The Influence of Design Type

With this background, it was possible to compute mean effect sizes for all variables that were theoretically relevant and represented by an adequate number of comparisons in our pool of research studies. For example, Table 9 reports the mean effect sizes for the two types of research designs included here, showing that pretest versus posttest comparisons report larger gains than studies that compare treatment groups to control groups.

Table 9

Mean Effect Sizes for Research Design Types

Research design type	*K*	Mean (*d*)	*SD* (*d*)	95% confidence interval
Pretest vs. posttest	44	.98	.92	.70 to 1.26
Treatment vs. control groups	44	.53	.82	.28 to .78

The greater gains found in pretest versus posttest designs has two major sources. First, it can be attributed in part to the inflation in estimated effect size for dependent sample designs resulting from using the *t* score divided by the square root of *N*, rather than the more accurate formula based on the standard deviations for Time 1 and Time 2.

However, the difference can also be attributed to the natural development in writing proficiency that comes with time. Further, since these pretest versus posttest designs usually consist of students working on multiple drafts of the same essay, the gains reflect the natural process of improvement that results from revision. Thus, the gains in performance in pre- versus posttest comparisons are influenced both by the feedback treatment and by the natural development that occurs over time in association with the revision process. This interpretation is further supported by two of these studies that reported pre-post test results for a control group that received no feedback. Thus, Hillocks (1982) reported an effect size of .82 for writing quality improvement with no feedback, and Brakel-Olson (1990) reported an effect size of 1.04 for holistic writing improvement with no feedback. These two comparisons were excluded from the meta-analysis, because no feedback was provided to students. However, they indicate that the particular pre-post test designs included in this study, many of which lacked control groups, can tell relatively little about the influence of feedback, because that treatment is confounded with the natural development that comes with time.

In contrast, the influence of natural development is accounted for in treatment-control group designs, because both groups practice writing for the same amount of time. We still see positive gains in writing proficiency in treatment-control designs, but the effect sizes are more modest. This latter finding answers the overall question: Does feedback on student writing, considered on its own, result in gains in writing development? The answer is yes, but the gains overall are not strong.

Because treatment-control design studies have greater experimental validity than pre-post design studies that include only a single treatment group (with no control group), we have given greater weight to the findings from those studies in our meta-analysis. Thus, for each of the following analyses, we report two sets of findings: 1) the mean effect sizes for the pretest-posttest designs and 2) the mean effect sizes for the experimental treatment-control designs. Further, when possible we present separate findings for L1-English students and for L2 students. However, some comparisons are based on too few effect sizes to permit this level of detail.

5.3. L1-English Versus L2 Groups

Table 10 shows that learners of a second/foreign language made larger gains associated with feedback than native speakers of English (in studies with treatment-control designs). This can in part be explained by the different outcome measures used in L1-English versus L2 studies: The former measured improvement in content scores or in holistic measures of writing quality, while the latter often measured improvement in grammatical accuracy (see Table 22 in Section 5.6). In addition, this difference might relate to proficiency level, since L1-English writers presumably are at a higher proficiency level than L2 students. That possibility is supported by the findings reported for the L2 group in Table 11, which shows that low-level students achieve larger gains in writing development than more advanced students. Apparently larger gains in proficiency are possible for low-level groups, simply because they need to learn so much. This pattern holds for low-proficiency compared to high-proficiency L2 students and apparently also holds for L2 students compared to L1 students. (There is little difference if we consider only the studies that employed pretest-posttest designs, shown especially by the overlaps in the confidence intervals.)

It was not possible to compare gains across age groups or across writing tasks, because many studies did not specify these characteristics, and for those that did, most studies relied on a single age group (university students) and a single task type (the pedagogical essay).

Table 10

Mean Effect Sizes for Each Language Group

Language group	*k*	Mean (*d*)	*SD* (*d*)	95% confidence interval
			Pretest/posttest design	
L1-English	8	1.20	.83	.50 to 1.89
L2-English	23	.92	.84	.56 to 1.28
L2-other	7	1.53	1.29	.33 to 2.72
			Treatment/control design	
L1-English	12	-0.03	.63	-.43 to .37
L2-English	24	0.66	.84	.30 to 1.01
L2-other	7	1.09	.55	.58 to 1.60

Table 11

Mean Effect Sizes for Language Proficiency Levels (L2 Students Only)

Language proficiency level	*k*	Mean (*d*)	*SD (d)*	95% confidence interval
			Pretest/posttest design	
Low/beginning	16	1.35	1.03	.80 to 1.90
High/advanced	17	1.03	.75	.64 to 1.42
			Treatment/control design	
Low/beginning	12	0.98	.94	.39 to 1.58
High/advanced	16	0.46	.87	.00 to .93

5.4. Source and Mode of Feedback

Turning to the different ways in which feedback was provided, we find several interesting patterns. On first consideration, there appears to be little overall difference between feedback provided by the teacher and feedback provided from other sources, as shown in Table 12.

However, further exploration of this general pattern shows that it hides a relatively large difference between L1-English and L2-English learners: As Table 13 shows, L1-English writers had much larger gains resulting from teacher feedback than from other feedback (peer or computer). In contrast, L2-English writers showed exactly the opposite trend, with much larger gains resulting from other feedback. (The same trends are shown for both design types, although some of the comparisons are based on only a few effect sizes.)

Table 12

Mean Effect Sizes for Different Sources of Feedback

Source	*k*	Mean (*d*)	*SD (d)*	95% confidence interval
			Pretest/posttest design	
Teacher	40	.96	.91	.67 to 1.25
Other	4	.19	n/a	n/a
		.31		
		1.96		
		2.33		
			Treatment/control design	
Teacher	25	.53	.60	.28 to .78
Other	19	.52	1.06	.01 to 1.04

Table 13

Mean Effect Sizes for Different Sources of Feedback—L1 English Versus L2 English

Feedback source	*k*	Mean *(d)*	*SD (d)*	95% confidence interval
		Pretest/post-test design for L1 English		
Teacher	6	1.52	.70	.78 to 2.25
Other	2	.19	n/a	n/a
		.31		
		Treatment/control design for L1 English		
Teacher	2	.53	n/a	n/a
		.53		
Other	10	-.14	.63	-.59 to .31
		Pretest/posttest design for L2 English		
Teacher	21	.80	.78	.45 to 1.16
Other	2	1.96	n/a	n/a
		2.33		
		Treatment/control design for L2 English		
Teacher	16	.28	.50	.02 to .55
Other	8	1.41	.92	.63 to 2.18

Surprisingly, oral feedback seems to have been more effective in these studies than written feedback, as shown in Table 14. (Again the same trends are shown for both design types.) However, similar to the findings on source of feedback, Table 15 shows that L1-English students differ from L2-English students in their preferred mode of feedback. Although based on only a few effect sizes, Table 15 shows that L1-English students achieved strong gains in writing proficiency resulting from oral feedback, contrasted with no or small gains resulting from written feedback. In contrast, L2-English students achieved moderately strong gains in writing proficiency following both oral and written feedback. These findings are based on relatively few effect sizes and so must be interpreted with caution. However, coupled with the findings on preferred source of feedback (Tables 12 and 13), they suggest an interesting difference in the typical learning styles of L1-English versus L2-English students.

Table 14

Mean Effect Sizes for Different Modes of Delivery of Feedback

Feedback mode	*k*	Mean (*d*)	*SD (d)*	95% confidence interval
			Pretest/posttest design	
Written	30	.68	.70	.42 to .94
Oral	12	1.86	.91	1.29 to 2.44
			Treatment/control design	
Written	30	.40	.91	.07 to .74
Oral	13	.84	.52	.52 to 1.15

Table 15

Mean Effect Sizes for Feedback Modes of Delivery—L1 English Versus L2 English

Feedback mode	*k*	Mean (*d*)	*SD (d)*	95% confidence interval
			Pretest/posttest design for L1 English	
Written	2	.61	n/a	n/a
		.81		
Oral	4	1.54	n/a	n/a
		1.80		
		1.86		
		2.47		
			Treatment/control design for L1 English	
Written	10	-.14	.63	-.59 to .31
Oral	2	.53	n/a	n/a
		.53		
			Pretest/post-test design for L2 English	
Written	21	.80	.78	.45 to 1.16
Oral	2	1.96	n/a	n/a
		2.33		
			Treatment/control design for L2 English	
Written	19	.69	.94	.24 to 1.14
Oral	5	.53	.35	.10 to .97

5.5. The Focus and Type of Feedback

Most studies in our sample provided feedback that focused on both content and form, but some studies focused strictly on one or the other. However, as Table 16 shows, feedback that focuses purely on form is less effective than feedback that focuses on content plus form. This finding seems to support the claim that writing tasks and feedback should be meaningful for students, with tasks that focus on the communication of information. Such communicative tasks are apparently very effective when coupled with a focus on form, while an exclusive focus on form (with no attention to content) is considerably less effective.

Table 16

Mean Effect Sizes for the Different Focuses of Feedback

Focus	*k*	Mean (*d*)	*SD (d)*	95% confidence interval
			Pretest/posttest design	
Content	2	.09		n/a
		1.22		
Form	16	.56	.55	.27 to .85
Content + form	24	1.20	1.02	.76 to 1.63
			Treatment/control design	
Content	2	.10		n/a
		1.12		
Form	8	.08	.44	-.28 to .45
Content + form	27	.48	.87	.14 to .82

This same pattern holds when we consider only L2-English learners. The results for treatment-control designs in Table 17 show that the greatest gains in writing development for L2-English writers were made with feedback that focused on content and form, while feedback focused on form resulted in no significant gains for this group. (The results for pretest-posttest designs show little difference here, especially when the overlap in confidence intervals is considered.)

Table 17

Mean Effect Sizes for the Different Focuses of Feedback (L2 English Only)

Focus	*k*	Mean (*d*)	*SD (d)*	95% confidence interval
			Pretest/posttest design	
Content	2	.09	n/a	n/a
		1.22		
Form	7	.91	.65	.31 to 1.51
Content + form	12	.77	.90	.20 to 1.34
			Treatment/control design	
Content	2	.10	n/a	n/a
		1.12		
Form	7	.03	.44	-.38 to .44
Content + form	10	.71	.88	.08 to 1.34

The findings on the different types of feedback, shown in Table 18, are consistent with the findings on feedback focus: commenting results in greater gains than error location.

These findings were somewhat difficult to interpret because many studies were vague in their descriptions of feedback type. It was often unclear what kinds of feedback were provided as comments, and many studies were mixed in that they both identified the location of some errors and provided comments in the margins and at the end of the paper. In addition, eight studies provided feedback in the form of general training (e.g., on the revision process) rather than specific feedback on a writing sample. Thus, Table 18 indicates that commenting is the most effective type of feedback, but the differences are small in the treatment-control studies.

Table 18

Mean Effect Sizes for the Different Types of Feedback

Type	*k*	Mean (*d*)	*SD (d)*	95% confidence interval
			Pretest/posttest design	
Comments	16	1.57	.96	1.06 to 2.08
Error location	10	.56	.92	-.10 to 1.22
Mixed	16	.75	.60	.43 to 1.08
Other	2	.19		n/a
		.31		
			Treatment/control design	
Comments	26	.40	.84	.06 to .74
Error location	4	.05	n/a	n/a
		.28		
		.46		
		2.28		
Mixed	7	.18	.21	-.02 to .37
Other	6	1.23	.80	.40 to 2.07

5.6. Comparing Different Types of Outcome Measures: The Different Ways in Which Writing Proficiency Can Develop

Finally, we can ask whether all aspects of writing development can be improved through feedback. This question can be investigated by comparing the effect sizes for the different kinds of outcome measures. Table 19 shows the comparison between the two general outcome types represented in these studies: scores of writing proficiency (reflecting accuracy or quality) versus a measure of the extent to which writing had been revised across multiple drafts. Gains were reported for both kinds of outcome measures, although those gains were much larger for measures of revising (regardless of quality or accuracy). Although based on only three effect sizes, this finding seems uncontroversial—and uninteresting: Students will make more revisions when they are given feedback that tells them that they should make revisions.

Table 19

Mean Effect Sizes for the Different Outcome Measures of Writing Development

Outcome measure	*k*	Mean (*d*)	*SD (d)*	95% confidence interval
			Pretest/posttest design	
Writing proficiency	44	.98	.92	.70 to 1.26
Revisions	1	.53	n/a	n/a
			Treatment/control design	
Writing proficiency	40	.42	.77	.18 to .67
Revisions	3	1.51	n/a	n/a
		1.90		
		2.25		

More insightful analyses are possible by considering only studies of improvement in writing proficiency, comparing the specific focuses of the outcome measure (e.g., a focus on grammar/form, spelling, content, or overall holistic quality). As Table 20 shows, gains were reported for most outcome measures. For measures of content, those gains were small and not significant (shown by the confidence interval including 0.0), but gains were larger for outcome measures focused on grammar and holistic measures of writing quality (which presumably reflects both form and content). (The gains for holistic quality were smaller for treatment-control designs.) In contrast, outcome measures of spelling accuracy actually showed a decrease following feedback.

When the language groups are distinguished, as in Table 21, we see that L2 students made gains in grammar/form and overall quality (and nonsignificant gains in content), but that the gains for L1 students were restricted to holistic ratings of overall quality.

Table 20

Mean Effect Sizes for the Different Focuses of Outcome Measures

Outcome focus	*k*	Mean (*d*)	*SD (d)*	95% confidence interval
			Pretest/posttest design	
Grammar/form	18	1.12	1.08	.58 to 1.66
Content	8	.40	.60	-.09 to .90
Holistic rating of quality	18	1.10	.79	.71 to 1.49
			Treatment/control design	
Grammar/form	18	.80	.78	.41 to 1.18
Content	11	.58	1.00	-.09 to 1.25
Holistic rating of quality	9	.51	.40	.20 to .81
Spelling	5	-.53	.33	-.95 to -.12

Table 21

Mean Effect Sizes for Different Focuses of Outcome Measures—L1 English Versus L2 English

Outcome focus	*k*	Mean (*d*)	*SD (d)*	95% confidence interval
			Pretest/posttest design for L1 English	
Holistic rating of quality	8	1.20	.83	.50 to 1.89
			Treatment/control design for L1 English	
Content	5	.24	.64	-.55 to 1.03
Spelling	5	-.53	.33	-.95 to -.12
			Pretest/posttest design for L2 English	
Holistic rating of quality	8	1.23	.74	.61 to 1.85
Content	6	.51	.67	-.20 to 1.21
Grammar/form	16	1.19	1.13	.59 to 1.79
			Treatment/control design for L2 English	
Holistic rating of quality	7	.56	.43	.16 to .96
Content	6	.86	1.22	-.42 to 2.14
Grammar/form	18	.80	.78	.41 to 1.18

5.7. Are Particular Kinds of Feedback Associated With Particular Gains in Writing Development?

The obvious question at this point is whether the particular type or focus of feedback results in specialized gains in writing development, reflected by the focus of the outcome measure. Table 22 shows the influence of feedback focus on different aspects of writing development.

The findings are not encouraging for the role of feedback in improving content ratings (considered only in treatment-control designs): Apparently students do not become more informative, logical, or elaborated in their prose as a result of feedback, regardless of the focus of that feedback. Thus, regardless of the focus of feedback, students made no significant gains in content scores.

Table 22

Mean Effect Sizes for Different Outcome Focuses, Depending on the Feedback Focus

Outcome focus	*k*	Mean (*d*)	*SD (d)*	95% confidence interval
			Pretest/posttest design	
Holistic quality rating				
Form feedback	8	.48	.41	.14 to .82
Content + form feedback	8	1.46	.65	.92 to 2.00
General feedback (unspecified focus)	2	1.96 2.33	n/a	n/a
Grammar/form accuracy rating				
Content + form feedback	12	1.36	1.18	.61 to 2.10
Form feedback	5	.77	.77	-.19 to 1.73
			Treatment/control design	
Content rating				
Content + form feedback	6	.25	.57	-.34 to .85
Grammar/form accuracy rating				
Content + form feedback	13	1.02	.81	.53 to 1.51
Form feedback	5	.22	.19	-.03 to .46

The overall holistic quality of an essay is probably the least informative of the outcome measures because it is impossible to identify the specific aspects of writing development that have improved. However, this outcome measure is also the most popular in pretest-posttest studies. Table 22 suggests that it is possible to achieve large gains in the holistic quality rating, resulting either from comments on content and form or from general feedback with an unspecified focus (which presumably includes comments). In contrast, feedback on form results in only small gains in these holistic quality scores.

Table 22 further shows that it is possible to improve grammatical accuracy. But the relative importance of the predictor variables is surprising here: Feedback focused on a combination of form and content results in a much greater improvement of grammatical accuracy than feedback that focuses exclusively on form. This difference is found for both design types. This finding suggests that student writing for real-world purposes, with the goal of communicating particular content, enables and encourages students to achieve greater gains in writing development than artificial writing tasks that are focused primarily on grammatical accuracy.

Finally, we can consider this same general question in relation to the different types of feedback. Although some of the comparisons are based on very few effect sizes, Table 23 suggests that specific feedback of any type is not very helpful for improving content ratings. In contrast, general training in the revision process does seem to help students improve the content of their papers. The results from pretest-posttest studies indicated that holistic quality ratings can be improved considerably by feedback in the form of comments, while specific comments tied to a particular location in the text provide less benefit for improvement in holistic quality.

The most interesting pattern in Table 23 has to do with the grammatical accuracy ratings, where the greatest improvements are associated with feedback in the form of comments. In contrast, error location feedback (either with or without more detailed comments) results in only small gains. Both design types show the same trend here. This finding applies only to L2 students (see Table 24), because none of these studies investigated grammatical accuracy for L1-English students. The patterns for gains in grammatical accuracy shown in Tables 22–24 are interesting because they suggest that students benefit more from general explanations of a grammatical phenomenon than from identification of specific errors. In fact, error identification seems to detract from the benefit of commenting: Students made smaller gains when feedback included

error identification, even if comments supplemented the identification of errors. Apparently, explanations of error patterns are more helpful than identifying selected specific errors.

Table 23

Mean Effect Sizes for Different Outcome Focuses, Depending on the Feedback Type

Outcome focus	*k*	Mean (*d*)	*SD (d)*	95% confidence interval
			Pretest/posttest design	
Content rating				
Feedback with error location	7	.29	.54	-.21 to .78
Holistic quality rating				
Feedback as comments	8	1.67	.66	1.12 to 2.23
Feedback with error location	10	.64	.55	.25 to 1.04
Grammar/form accuracy rating				
Feedback as comments	7	1.50	1.33	.27 to 2.73
Feedback with error location	11	.88	.88	.29 to 1.47
			Treatment/control design	
Content rating				
Feedback as comments	8	.09	.63	-.44 to .62
Training on revision process	3	1.51 1.90 2.25	n/a	n/a
Grammar/form accuracy rating				
Feedback as comments	8	1.18	.73	.57 to 1.79
Feedback with error location	10	.49	.70	-.02 to 1.32

Table 24

Mean Effect Sizes for Different Outcome Focuses, Depending on the Feedback Type (L2 Students Only)

Outcome focus	*k*	Mean (*d*)	*SD (d)*	95% confidence interval
			Pretest/posttest design	
Grammar/form accuracy rating				
Feedback as comments	7	1.50	1.33	.27 to 2.73
Feedback with error Location	9	.95	.96	.21 to 1.69
			Treatment/control design	
Grammar/form accuracy rating				
Feedback as comments	8	1.18	.73	.57 to 1.79
Feedback with error location	10	.49	.70	-.02 to .99

6. Discussion and Implications for TOEFL

Several important general patterns emerge from this synthesis of research on the effectiveness of feedback for individual writing development:

1. Interest in this research question is widespread: More than 300 studies have been published on this topic in the last 25 years.

2. But the large majority of studies in this research domain are not suitable for inclusion in a meta-analysis: Less than 10% of the studies in our sample were found to be suitable. There were three general reasons for this:

   2a. Many studies in this domain were qualitative, or thought pieces.

   2b.Many of the quantitative studies in this domain were not carefully designed, or the reporting standards were not adequate for the purposes of meta-analysis.

   2c. Some quantitative studies addressed different research questions from the one that we are focusing on here.

3. Both L1-English and L2-English students make gains in writing development in response to feedback.

4. But lower proficiency levels make greater gains in writing development in response to feedback than students with high proficiency levels: L2 students make greater gains than L1 students, and low proficiency L2 students make greater gains than high proficiency L2 students.

5. Large differences exist in how L1-English students and L2-English students respond to feedback from different sources and different modes.

   5a The greatest gains for L1-English students are achieved in response to teacher feedback presented orally.

   5b. The greatest gains for L2-English students are achieved in response to other feedback, including feedback from other students and feedback from computer.

6. A combined focus on content + form results in greater gains in writing development than an exclusive focus on form.

   6a. If we consider only L2-English learners, the differential influence of a focus on content + form versus form becomes even greater.

7. Larger gains in writing development result from feedback that is expressed through comments than from locating/correcting errors.

   7a These patterns are stronger for L2-English students, with moderately large gains resulting from comments, versus smaller or insignificant gains from error location feedback, and only small gains resulting from other feedback (including direct error correction and training in the revision process).

8. It is apparently difficult to provide specific feedback that improves the content of student writing.

   8a. Providing specific feedback on previous writing samples, whether through comments or through identifying specific trouble spots in the paper, results in only small improvements in content scores.

   8b. In contrast, providing training in the revision process results in large gains in content scores (based on only three effect sizes).

9. Providing feedback expressed through comments with a combined focus on content + form improves holistic quality ratings.

    9a. Error location or an exclusive focus on form results in only small gains for holistic quality.

10. Grammatical accuracy can be best improved by feedback that focuses on a combination of form and content

    10a. Feedback that focuses exclusively on form does not result in a significant improvement in grammatical accuracy

    10b. Feedback provided as written comments results in large gains in grammatical accuracy

    10c. Feedback as error location results in smaller or no gains in grammatical accuracy.

Some of these findings are surprising, running counter to our prior expectations. For example, a widely held perception notes that teachers have more authority and prestige in many nonwestern cultures than in American and British society. Because of that, we expected that teacher feedback would be more influential for L2-English students than for L1-English students. Previous studies on peer editing have found that L1-English students are receptive to this approach and make modest gains (Graham & Perin, 2007). In contrast, previous studies of L2-English students noted skepticism of the value of feedback from other students and that greater gains are made from teacher feedback than from peer feedback (see, e.g., Tsui & Ng, 2000). Thus, we expected that teacher feedback would be more influential than peer or computer feedback for L2-English students. However, those predictions were not supported by the meta-analysis. Rather, L1-English students showed strong gains from teacher feedback and no gains from peer feedback, while L2-English students showed greater gains from peer or computer feedback than from teacher feedback. The findings on mode of delivery are also surprising: L1-English students showed strong gains from oral feedback and no overall gains from written feedback, while L2-English students showed moderately strong gains from both oral and written feedback.

These findings further indicate that L2-English students are more adaptive than L1-English students. That is, L2-English students make gains in writing proficiency regardless of how feedback is presented: oral or written, from teachers or peers. This finding could in part be a

reflection of their lower proficiency status, but it also seems to reflect their ability to benefit from many different types of feedback. In contrast, L1-English students are quite polarized in terms of the kinds of feedback that are effective: From the teacher, presented orally; peer feedback and feedback presented in writing actually resulted in a loss of writing proficiency for L1 writers in these studies.

Similarly, the findings on the focus of feedback are noteworthy: A combined focus on content + form is generally much more effective than an exclusive focus on form for writing development. This trend is stronger for L2-English students than for L1-English students.

This pattern holds for outcomes that measure the overall holistic quality of student writing. But even for the improvement of grammatical accuracy, combined feedback on content plus form is more effective than feedback focused exclusively on form, supporting the general claims of Truscott (2007). (Unfortunately, no study in our sample investigated the influence of feedback focused exclusively on content for the development of grammatical accuracy.) These patterns seem to support the general approach advocated by proponents of content-based instruction (CBI), showing that real-world tasks with a focus on communicating actual content are more effective learning environments than tasks focused exclusively on grammatical form.

A similar trend seems to be at work for outcomes that measure improvement in content scores. In this case, specific feedback provided on previous written papers did not result in a significant improvement in content scores. In contrast, training in the revision process itself resulted in large improvements in content scores. This finding could have the same explanation as above: Helping students learn how to revise, with a focus on effective communication rather than a specific written product, results in the largest gains in proficiency (at least for the content of writing).

In contrast, specific feedback provided on papers was effective for improvements in holistic quality and in grammatical accuracy. In these cases, it was surprising that the type of feedback seemed equally important to the focus of feedback. It was predictable that feedback in the form of written comments would result in greater gains in overall holistic quality than feedback that identified the location of errors. However, the surprising findings here have to do with the improvement of grammatical accuracy: Feedback provided through written comments was found to be more effective for improving grammatical accuracy than error location. This finding suggests that students benefit the most from descriptions and/or explanations of their

grammatical patterns. In contrast, students might regard direct error identification as simple editing corrections, and so they seem less likely to generalize from those corrections to other instances of similar constructions. Here again, these findings are consistent with Truscott's (2007) findings that students improve little in their grammatical accuracy based on direct error correction.

All of the above conclusions should be treated with caution, because the meta-analysis has major limitations. First of all, this research domain was not very "mature" when evaluated for the purposes of meta-analysis. Although we were able to identify 306 published research studies that investigated the effectiveness of feedback on student writing, only 25 of those studies (less than 8%) proved to be suitable for inclusion in the statistical meta-analysis. Because a sample size of 25 is too small to permit comparisons for many of the parameters of interest, we permitted the inclusion of multiple effect sizes from the individual studies. But this decision introduced the risk that a single aberrant study (e.g., with a flawed design or methods) might have a relatively large influence on the overall results of the meta-analysis. Finally, even with this compromise, we ended up with only 88 effect sizes, and as a result, several of the specific findings from the meta-analysis were based on a sample of fewer than five effect sizes. In particular, the small sample size restricted the extent to which we could examine interactions among variables, and as a result, the influence of some variables could be confounded.

In addition, meta-analysis is inherently reductive in nature, and as a result, many of the particular findings from individual studies are discounted. To be published, it is usually necessary for a study to be innovative, filling some gap in the previous literature. Thus, direct replications of previous research are almost never published. And as a result, none of these studies are exactly comparable. For the purposes of the meta-analysis, we collapsed numerous more specific measures into a few general categories. But in fact, studies employed many different specific treatments and measured many different specific outcomes. This is an important caveat that should be applied to any quantitative meta-analysis: Its strength is identifying the general trends that hold across a research domain, but its major weakness is that those generalizations obscure the individual patterns of variation found across studies.

Thus, all of the general findings described in this report should be subjected to further research, with more tightly controlled designs. However, the results are interesting, with three general findings especially worthy of future research:

1. L2-English students seem to be very receptive to feedback from sources other than teachers.
2. Feedback on content is at least as important as feedback on form. Even when the writing development goals are to improve grammatical accuracy, feedback on form coupled with feedback on content is more effective than feedback focused exclusively on form.
3. Feedback in the form of written comments is more effective than simple error location, again even when the writing development goals are to improve grammatical accuracy.

References

Ashwell, T. (2000). Patterns of teacher response to student writing in a multiple-draft composition classroom: Is content feedback followed by form feedback the best method? *Journal of Second Language Writing, 9*, 227–257.

Becker, L. A. (1999). *Effect size calculator* [Computer software]. Available at http://web.uccs.edu/lbecker/Psy590/escalc3.htm

Berg, E. C. (1999). The effects of trained peer response ESL students' revision types and writing quality. *Journal of Second Language Writing, 9*, 215–241.

Bitchner, J., Young, S., & Cameron, D. (2005). The effect of different types of corrective feedback on ESL student writing. *Journal of Second Language Writing, 14*, 191–205.

Brakel-Olson, V. L. (1990). The revising processes of sixth-grade writers with and without peer feedback. *Journal of Educational Research, 84*, 22–29.

Chandler, J. (2003). The efficacy of various kinds of error correction for improvement of the accuracy and fluency of L2 student writing. *Journal of Second Language Writing, 12*, 267–296.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

DiPardo, A., & Freedman, S. W. (1988). Peer response groups in the writing classroom: Theoretical foundations and new directions. *Review of Educational Research, 58*, 119–149.

Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods, 1*, 170–177.

Ferris, D. (1995). Student reactions to teacher response in multiple-draft composition classrooms. *TESOL Quarterly, 29*, 33–53.

Ferris, D. (1999). The case for grammar correction in L2 writing classes: A response to Truscott (1996). *Journal of Second Language Writing, 8*, 1–10.

Ferris, D. (2003). *Response to student writing*. Mahwah, NJ: Lawrence Erlbaum Associates.

Ferris, D. (2006). Does error feedback help student writers? New evidence on the short- and long-term effects of written error correction. In K. Hyland & F. Hyland (Eds.), *Feedback*

in second language writing: Contexts and issues (pp. 81–104). Cambridge, UK: Cambridge University Press.

Ferris, D., & Roberts, B. (2001). Error feedback in L2 writing classes: How explicit does it need to be? *Journal of Second Language Writing, 10*, 161–184.

Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology, 99,* 445–476.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Hillocks, G. (1982). The interaction of instruction, teacher comment, and revision in teaching the composing process. *Research in the Teaching of English, 16*, 261–277.

Hyland, F. (2003). Focusing on form: Student engagement with teacher feedback. *System, 31*, 217–230.

Hyland, K., & Hyland, F. (2006). Interpersonal aspects of response: Constructing and interpreting teacher written feedback. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 206–224). Cambridge, UK: Cambridge University Press.

Leki, I. (2006). "You cannot ignore": Graduate L2 students' experience of and response to written feedback practices within their disciplines. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 266–285). Cambridge, UK: Cambridge University Press.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

McGroarty, M. E., & Zhu, W. (1997). Triangulation in classroom research: A study of peer revision. *Language Learning, 47,* 1–43.

Min, H.-T. (2006). The effects of trained peer review on EFL students' revision types and writing quality. *Journal of Second Language Writing, 15*, 118–141.

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning, 50*, 417–528.

Norris, J. M., & Ortega, L. (2006). *Synthesizing research on language learning and teaching*. Philadelphia, PA: John Benjamins.

Norris, J. M., & Ortega, L. (2007). The future of research synthesis in applied linguistics: Beyond art or science. *TESOL Quarterly, 41*, 805–815.

Robb, T., Ross, S., & Shortreed, I. (1986). Salience of feedback on error and its effect on EFL writing quality. *TESOL Quarterly, 20*, 83–93.

Rosenthal, R., Rosnow, R. L., & Rubin, D. B. (2000). *Contrasts and effect sizes in behavioral research*. Cambridge, UK: Cambridge University Press.

Truscott, J. (2007). The effect of error correction on learners' ability to write accurately. *Journal of Second Language Writing, 16*, 255–272.

Tsui, A. B. M., & Ng, M. (2000). Do secondary L2 writers benefit from peer comments? *Journal of Second Language Writing, 9*, 147–170.

Woods, A., Fletcher, P., & Hughes, A. (1986). *Statistics in language studies*. Cambridge, UK: Cambridge University Press.

Notes

[1] Random assignment of subjects into treatment or control groups was not practiced in nearly all of these studies and was thus not applied as a criterion for exclusion.

[2] In some large-scale meta-analyses, the estimates of each individual effect size are weighted according to the sample size of the study, either directly or by using *inverse variance weights*. Then, the study reports a *weighted mean effect size* rather than a simple arithmetic mean of effect sizes. (See, e.g., Hedges & Olkin, 1985; Lipsey & Wilson, 2001.) However, this practice has generally not been adopted in previous meta-analyses in applied linguistics (e.g., Norris & Ortega, 2000; Truscott, 2007), and so it was not employed in the present analysis.

Appendix A

Studies Used for the Research Synthesis

References marked with an asterisk indicate studies included in the meta-analysis.

Allal, L., Lopez, L. M., Lehraus, K., & Forget, A. (2005). Whole-class and peer interaction in an activity of writing and revision. In T. Kostouli (Ed.), *Writing in context(s): Textual practices and writing processes in sociocultural settings* (pp. 69–91). New York, NY: Springer.

Allwright, R. L., Woodley, M-P., & Allwright, J. M. (1988). Investigating reformulation as a practical strategy for the teaching of academic writing. *Applied Linguistics, 9,* 236–256.

Anderson, K., Benson, C., & Lynch, T. (2001). *Feedback on writing: Attitudes and uptake*. Retrieved from ERIC database. (ED463672)

Andrade, H. & Boulay, B. (2003). Role of rubric-referenced self-assessment in learning to write. *The Journal of Educational Research, 97*(1), 21–34.

Anson, C. M. (2000). Response and the social construction of error. *Assessing Writing, 7,* 5–21.

Artemeva, A., & Logie, S. (2002). Engineering students to intellectual teamwork: The teaching and practice of peer feedback in the professional communication classroom. *Language and Learning Across the Disciplines, 6*(1). Retrieved from http://wac.colostate.edu/llad/v6n1/artemeva.pdf

*Ashwell, T. (2000). Patterns of teacher response to student writing in a multiple-draft composition classroom: Is content feedback followed by form feedback the best method? *Journal of Second Language Writing, 9,* 227–257.

Atari, O. F., & Triki, M. A. (2000). The formal features of oral and literate strategies of communication: Their implications for EFL writing revision. *International Review of Applied Linguistics, 38,* 95–107.

Bailey, J., & Vardi, I. (1999). Iterative feedback: Impacts on student writing. Paper presented at the HERDSA Annual International Conference, Melbourne, Australia.

Bardine, B., Bardine, M., & Deegan, E. (2000). Beyond the red pen: Clarifying our role in the response process. *English Journal, 90*(1), 94–101.

Beach, R. (1976). Self-evaluation strategies of extensive revisers and non-revisers. *College Composition and Communication, 27,* 160–164.

Beach, R. (1979). The effects of between-draft teacher evaluation versus student self-evaluation on high school students' revising of rough drafts. *Research in the Teaching of English, 13*, 111–119.

Beason, L. (1993). Feedback and revision in writing across the curriculum classes. *Research in the Teaching of English, 27*, 395–421.

Belcher, D. (1990). Peer vs. teacher response in the advanced composition class. *Issues in Writing, 2*, 128–150.

Belcher, D. (1994). The apprenticeship approach to advanced academic literacy: Graduate students and their mentors. *English for Specific Purposes, 13*, 23–34.

*Berg, E. C. (1999). The effects of trained peer response ESL students' revision types and writing quality. *Journal of Second Language Writing, 9*, 215–241.

Bernhardt, S. A. (1988). Text revisions by basic writers: Impromptu first draft to take-home revision. *Research in the Teaching of English, 22*, 266–280.

*Bitchner, J., Young, S., & Cameron, D. (2005). The effect of different types of corrective feedback on ESL student writing. *Journal of Second Language Writing, 14*, 191–205.

Blain, S. (2001). Study of verbal peer feedback on the improvement of the quality of writing and the transfer of knowledge in francophone students in grade 4 living in a minority situation in Canada. *Language, Culture, and Curriculum, 14*(1), 156–170.

Bolt, P. (1992). An evaluation of grammar-checking programs as self-help learning aids for learners of English as a foreign language. *Computer Assisted Language Learning, 5*, 49–91.

Braine, G. (1997). Beyond word processing: Networked computers in ESL writing classes. *Computers and Composition, 14*, 45–58.

*Brakel Olson, V. L. (1990). The revising processes of sixth-grade writers with and without peer feedback. *Journal of Educational Research, 84*, 22–29.

Brice, C. (1995). *ESL writers' reactions to teacher commentary: A case study*. Retrieved from ERIC database. (ED394312)

Bridwell, L. S. (1980). Revising strategies in twelfth grade students' transactional writing. *Research in the Teaching of English, 14*, 197–222.

Brown, A. (2005). Self-assessment of writing in independent language learning programs: The value of annotated samples. *Assessing Writing, 10*, 174–191.

Burston, J. (2001). Computer-mediated feedback in composition correction. *CALICO Journal, 19*(1), 37–50.

Caffarella, R., & Barnett, B. (2000). Teaching doctoral students to become scholarly writers: The importance of giving and receiving critiques. *Studies in Higher Education, 25*(1), 39–52.

Callahan, S. (2000). Responding to the invisible student. *Assessing Writing, 7*, 57–77.

*Cardelle, M., & Corno, L. (1981). Effects on second language learning of variations in written feedback on homework assignments. *TESOL Quarterly, 15*, 251–261.

Carson, J. G., & Nelson, G. L. (1994). Writing groups: Cross-cultural issues. *Journal of Second Language Writing, 3*, 17–30.

Carson, J. G., & Nelson, G. L. (1996). Chinese students' perceptions of ESL peer response group interaction. *Journal of Second Language Writing, 5*, 1–19.

Caulk, N. (1994). Comparing teacher and student responses to written work. *TESOL Quarterly, 28*, 181–187.

*Chandler, J. (2003). The efficacy of various kinds of error feedback for improvement in the accuracy and fluency of L2 student writing. *Journal of Second Language Writing, 12*, 267–296.

Chandler, J. (2004). A response to Truscott. *Journal of Second Language Writing, 13*, 345–348.

Chandrasegeran, A. (1986). An exploratory study of EFL student revision and self-correction skills. *RELC Journal*, *17*(2), 26–40.

Chandrasegeran, A., & Ellis, M. (2005). Essay assist—Developing software for writing skills improvement in partnership with students. *RELC Journal, 36*, 137–155.

Charles, M. (1990). Responding to problems in written English using a student self-monitoring technique. *ELT Journal*, *44*(4), 286–293.

Chaudron, C. (1984). The effects of feedback on students' composition revisions. *RELC Journal, 15*(2), 1–15.

Chen, J. F. (1997). Computer generated error feedback and writing process: A link. *TESL-EJ, 2*(3), 1–19.

Chi, F-M. (1999). *The writer, the teacher, and the text: examples from Taiwanese EFL college students*. Retrieved from ERIC database. (ED442272)

Chiu, C-Y., & Savignon, S. J. (2006). Writing to mean: Computer-mediated feedback in online tutoring of multidraft compositions. *CALICO Journal, 24*, 97–114.

Cho, K., & Schunn, C. D. (2007). Scaffolded writing and rewriting in the discipline: A web-based reciprocal peer review system. *Computers & Education, 48*. 409–426.

Cho, K., Schunn, C. D., & Charney, D. (2006). Commenting on writing: Typology and perceived helpfulness of comments from novice peer reviewers and subject matter experts. *Written Communication, 23*, 260–294.

Cho, K., Schunn, C. D., & Wilson, R. W. (2006). Validity and reliability of scaffolded peer assessment of writing from instructor and student perspectives. *Journal of Educational Psychology, 98*, 891–901.

Clare, L., Valdes, R., & Patthey-Chavez, G. (2000). *Learning to write in urban elementary and middle schools: An investigation of teachers' written feedback on student compositions*. Retrieved from ERIC database. (ED446207)

Cohen, A. D. (1987). Student processing of feedback on their compositions. In A. L. Wenden & J. Rubin (Eds.), *Learner strategies in language learning* (pp. 57–69). Englewood Cliffs, NJ: Prentice Hall.

Cohen, A. D., & Cavalcanti, M. C. (1990). Feedback on compositions: Teacher and student verbal reports. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 155–177). Cambridge, UK: Cambridge University Press.

Connor, U., & Asenavage, K. (1994). Peer response groups in ESL writing classes: How much impact on revision? *Journal of Second Language Writing, 3*, 257–276.

Connors, R. J., & Lunsford, A. (1988). Frequency of formal errors in current college writing, or Ma and Pa Kettle do research. *College Composition and Communication, 39*(4), 395–409.

Connors, R. J., & Lunsford, A. (1993). Teachers' rhetorical comments on student papers. *College Composition and Communication, 44*, 200-–23.

Crank, V. (2002). Asynchronous electronic peer response in a hybrid basic writing classroom. *Teaching English in the Two Year College, 30*, 145–155.

Crehan, K., & Curfman, M. (1999). *The effect on performance of rapid feedback on state writing assessments*. Retrieved from ERIC database. (ED435629)

Cresswell, A. (2000). Self-monitoring in student writing: Developing learner responsibility. *ELT Journal, 54*, 235–244.

Dalgish, G. (1991). Computer-assisted error analysis and courseware design: Applications for ESL in the Swedish context. *CALICO Journal, 9*(2), 39–56.

*Davis, W., & Fulton, J. (1997). *The effects of professors' feedback on the growth of students' overall writing quality in two college freshman English courses*. Retrieved from ERIC database. (ED414570)

*Davis, W., & Mahoney, K. (1999). *The effects of computer skills and feedback on the gains in students' overall writing quality in college freshman composition courses*. Retrieved from ERIC database. (ED435097)

De Guerrero, M. C. M., & Villamil, O. S. (1994). Social-cognitive dimensions of interaction in L2 peer revision. *Modern Language Journal, 78*, 484–496.

De Guerrero, M. C. M., & Villamil, O. S. (2000). Activating the ZPD: Scaffolding in L2 peer revision. *Modern Language Journal, 84*, 51–68.

Dheram, P. K. (1995). Feedback as a two-bullock cart: A case study of teaching writing. *ELT Journal, 49*(2), 160–168.

Diab, R. L. (2005). Teachers' and students' beliefs about responding to ESL writing: A case study. *TESL Canada Journal, 23*, 28–43.

DiGiovanni, E., & Nagaswami, E. (2001). Online peer review: An alternative to face-to-face? *ELT Journal, 55*, 263–272.

DiPardo, A., & Freedman, S. W. (1988). Peer response groups in the writing classroom: Theoretical foundations and new directions. *Review of Educational Research, 58*, 119–149

Dodigovic, M. (2002). Developing writing skills with a cyber-coach. *Computer Assisted Language Learning, 15*, 9–25.

Dohrer, G. (1991). Do teachers' comments on students' papers help? *College Teaching, 39*, 48–54.

Dwyer, H., & Sullivan, H. (1993). Student preferences for teacher and computer composition marking. *Journal of Educational Research, 86*(3), 137–141.

Enginarlar, H. (1993). Student response to teacher feedback in ESL writing. *System, 21*, 193–204.

Evans, K. A. (1997). Writing, response, and contexts of production or, Why it just wouldn't work to write about those bratty kids. *Language and Learning Across the Disciplines, 2*(1). Retrieved from http://wac.colostate.edu/llad/v2n1/evans.pdf

Eyres, S., Hatch, D., Turner, S., & West, M. (2001). Doctoral students' responses to writing critique: Message for teachers. *Journal of Nursing Education, 40*(4), 149–155.

Faigley, L., & Witte, E. (1981). Analyzing revision. *College Composition and Communication, 32*, 400–414.

*Fathman, A., & Whalley, E. (1990). Teacher response to student writing: Focus on form versus content. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 178–190). Cambridge, UK: Cambridge University Press.

Fazio, L. L. (2001). The effect of corrections and commentaries on the journal writing accuracy of minority- and majority-language students. *Journal of Second Language Writing, 10*, 235–249.

Ferris, D. R. (1995a). Can advanced ESL students be taught to correct their most serious and frequent errors? *CATESOL Journal, 8*(1), 41–62.

Ferris, D. R. (1995b). Student reactions to teacher response in multiple-draft composition classrooms. *TESOL Quarterly, 29*, 33–53.

Ferris, D. R. (1995c). Teaching ESL composition students to become independent self-editors. *TESOL Journal, 4*(4), 18–22.

Ferris, D. R. (1997). The influence of teacher commentary on student revision. *TESOL Quarterly, 31*, 315–339.

Ferris, D. R. (1999). The case for grammar correction in L2 writing classes: A response to Truscott (1996). *Journal of Second Language Writing, 8*, 1–11.

Ferris, D. R. (2001). Teaching writing for academic purposes. In Flowerdew, J., & Peacock, M. (Eds.), *Research perspectives on English for academic purposes* (pp. 298–314). Cambridge, UK: Cambridge University Press.

Ferris, D. R. (2004). The ''Grammar Correction'' debate in L2 Writing: Where are we, and where do we go from here? (and what do we do in the meantime . . .?). *Journal of Second Language Writing, 13*, 49–62.

*Ferris, D. (2006). Does error feedback help student writers? New evidence on the short- and long-term effects of written error correction. In K. Hyland & F. Hyland (Eds.), *Feedback*

in second language writing: Contexts and issues (pp. 81–105). Cambridge, UK: Cambridge University Press.

Ferris, D. R., Pezone, S., Tade, C. R., & Tinti, S. (1997). Teacher commentary on student writing: Descriptions & implications. *Journal of Second Language Writing, 6*, 155–182.

*Ferris, D. R., & Roberts, B. (2001). Error feedback in L2 writing classes: How explicit does it need to be? *Journal of Second Language Writing, 10*, 161–184.

Figueredo, L., & Varnhagen, C. K. (2006). Spelling and grammar checkers: Are they intrusive? *British Journal of Educational Technology, 37*, 721–732.

Fitzgerald, J., & Stamm, C. (1990). Effects of group conferences on first graders' revision in writing. *Written Communication, 7*, 96–135.

Flower, L., Hayes, J. R., Carey, L., Schriver, K., & Stratman, J. (1986). Detection, diagnosis, and the strategies of revision. *College Composition and Communication, 37*, 16–55.

Frankenberg-Garcia, A. (1999). Providing student writers with pre-text feedback. *ELT Journal, 53*, 100–106.

Frantzen, D. (1995). The effects of grammar supplementation on written accuracy in an intermediate Spanish content course. *Modern Language Journal, 79*, 329–344.

Frantzen, D., & Rissel, D. (1987). Learner self-correction of written compositions: What does it show us? In B. VanPatten, T. R. Dvorak, & J. F. Lee (Eds.), *Foreign language learning: A research perspective* (pp. 92–107). Cambridge, England: Newbury House.

Franzke, M., Kintsch, E., Caccamise, D., & Dooley, S. (2005). Summary street: Computer support for comprehension and writing. *Journal of Educational Computing Research, 35*(1), 53–80.

Gaskell, D., & Cobb, T. (2004). Can learners use concordance feedback for writing errors? *System, 32*, 301–319.

Goldstein, L. M. (2004). Questions and answers about teacher written commentary and student revision: Teachers and students working together. *Journal of Second Language Writing, 13*, 63–80.

Goldstein, L. M., & Conrad, S. M. (1990). Student input and negotiation of meaning in ESL writing conferences. *TESOL Quarterly, 24*, 443–460.

Goodfellow, R., Lamy, M-N., Jones, G. (2002). Assessing learners' writing using lexical frequency. *ReCALL, 14*, 133–145.

Goodrich Andrade, H., & Boulay, B. A. (2003). Role of rubric-referenced self-assessment in learning to write. *Journal of Educational Research, 97*, 21–34.

Greenman, C. (2004). Coaching academic English through voice and text production models, *ReCALL, 16*, 51–70.

Guénette, D. (2007). Is feedback pedagogically correct? Research design issues in studies of feedback on writing. *Journal of Second Language Writing, 16*, 40–53.

Hacker, D. J., Plumb, C., Butterfield, E. C., Quathamer, D., & Heineken, E. (1994). Text revision: Detection and correction of errors. *Journal of Educational Psychology, 86*, 65–78.

Hall, C. (1990). Managing the complexity of revising across language. *TESOL Quarterly, 24*, 43–60.

Hamid, M. O. (2007). Identifying second language errors: How plausible are plausible reconstructions? *ELT Journal, 61*, 107–116.

Hansen, J. G., & Liu, J. (2005). Guiding principles for effective peer response. *ELT Journal, 59*, 31–38.

Haswell, R. (2006). The complexities of responding to student writing; or, looking for shortcuts via the road of excess. *Across the Disciplines, 3*. Retrieved from http://wac.colostate.edu/atd/articles/haswell2006.cfm

Hayes, J. R., Flower, L., Schriver, K. A., Stratman, J. F., & Carey, L. (1987). Cognitive processes in revision. In S. Rosenberg (Ed.), *Advances in applied psycholinguistics: Vol. 2. Reading, writing, and language learning* (pp. 176–240). New York, NY: Cambridge University Press.

Hedgcock, J., & Lefkowitz, N. (1992). Collaborative oral/aural revision in foreign language writing instruction. *Journal of Second Language Writing, 1*, 255–276.

Hedgcock, J., & Lefkowitz, N. (1994). Feedback on feedback: Assessing learner receptivity to teacher response in L2 composing. *Journal of Second Language Writing, 3*, 141–163.

Hedgcock, J., & Lefkowitz, N. (1996). Some input on input: Two analyses of student response to expert feedback in L2 writing. *Modern Language Journal, 80*, 287–308.

Heift, T. (2001). Error-specific and individualised feedback in a Web-based language tutoring system: Do they read it? *ReCALL, 13*, 99–109.

Heift, T. (2004). Corrective feedback and learner uptake in CALL. *ReCALL, 16*(2), 416–431.

Heift, T. (2006). Context-sensitive help in CALL. *Computer Assisted Language Learning, 19,* 243–259.

Hendrickson, J. (1978). Error correction in foreign language teaching. Recent theory, research and practice. *Modern Language Journal, 80*, 387–398.

Hendrickson, J. (1980). The treatment of error in written work. *Modern Language Journal, 82*, 216–221.

Hewett, B. L. (2000). Characteristics of interactive oral and computer-mediated peer group talk and its influence on revision. *Computers and Composition, 17*, 265–288.

*Hillocks, G. (1982). The interaction of instruction, teacher comment, and revision in teaching the composing process. *Research in the Teaching of English, 16*, 261–277.

Holdich, C. E., Holdich, R. G., & Chung, P. W. H. (2002). Assessing aspects of children's written grammar: Automating the process. *Computers & Education, 39*, 37–50.

Hopkins, S., & Youngs, E. (1988). A FAIR response to student writing. *Computers and Composition, 5*, 83–97.

Horner, B. (1992). Rethinking the "sociality" of error: Teaching editing as negotiation. *Rhetoric Review, 11*(1), 172–199.

Huang, S-Y. (2000a). *The nature of an EFL teacher's audiotaped and written feedback on student writing: A case study*. Retrieved from ERIC database. (ED438728)

Huang, S-Y. (2000b). *A quantitative analysis of audiotaped and written feedback produced for students' writing and students' perceptions of the two feedback methods*. Retrieved from ERIC database. (ED448604)

Hunter, D., Mayenga, C., & Gambell, T. (2006). Classroom assessment tools and uses: Canadian English teachers' practices for writing. *Assessing Writing, 11*, 42–65.

Hyland, F. (1998). The impact of teacher written feedback on individual writers. *Journal of Second Language Writing, 7*, 255–286.

Hyland, F. (2000). ESL writers and feedback: giving more autonomy to students. *Language Teaching Research, 4*, 33–54.

Hyland, F. (2001a). Dealing with plagiarism when giving feedback. *ELT Journal, 55*, 375–381.

Hyland, F. (2001b). Teacher management of writing workshops: Two case studies. *Canadian Modern Language Review, 57*, 272–294.

Hyland, F. (2003). Focusing on form: Student engagement with teacher feedback. *System, 31*, 217–230.

Hyland, F., & Hyland, K. (2001). Sugaring the pill: Praise and criticism in written feedback. *Journal of Second Language Writing, 10*, 185–212.

Hyland, K. (1990). Providing productive feedback. *ELT Journal, 44*(4), 279–285.

Hyland, K., & Anan, E. (2006). Teachers' perceptions of error: The effects of first language and experience. *System, 34*, 509–519.

Hyland, K., & Hyland, F. (2006). Feedback on second language students' writing. *Language Teaching, 39*(2), 83–101.

Jacobs, G. (1987). First experiences with peer feedback on compositions: Student and teacher reaction. *System, 15*, 325–333.

Jacobs, G. (1989). Miscorrection in peer feedback in writing class. *RELC Journal, 20*(1), 68–75.

Jacobs, G. M., Curtis, A., Braine, G., & Huang, S.-Y. (1998). Feedback on student writing: Taking the middle path. *Journal of Second Language Writing, 7*, 307–317.

Jansen, C. J. M., Steehouder, M. F., Pilot, A., Schrauwen, D., & Looijmans, P. J. M. (1986). ALEXIS: Computer-assisted feedback on written assignments. *Computers and Composition, 4*(1), 32-–45.

Jeffery, F., & Selting, B. (1999). Reading the invisible ink: Assessing the responses of non-composition faculty. *Assessing Writing, 6*, 179–197.

Johanson, R. (1999). *Rethinking the red ink: audio-feedback in the ESL writing classroom*. Retrieved from ERIC database. (ED467865)

Johnson, D. M. (1992). Interpersonal involvement in discourse: Gender variation in L2 writers' complimenting strategies. *Journal of Second Language Writing, 1*, 195–215.

Jones, I. (1998). The effect of computer-generated spoken feedback on kindergarten students' written narratives. *Journal of Computing in Childhood Education 9*(1), 43–56.

*Kamimura, T. (2006). Effects of peer feedback on ESL student writers at different levels of proficiency: A Japanese context. *TESL Canada Journal, 23*(2), 12–39.

Kastman Breuch, L. M., & Racine, S. J. (2000). Developing sound tutor training for online writing centers: Creating productive peer reviewers. *Computers and Composition, 17*, 245–263.

Keh, C. L. (1990). Feedback in the writing process: A model and methods for implementation. *ELT Journal, 44*(4), 294–304.

Kepner, C. G. (1991). An experiment in the relationship of types of written feedback to the development of second-language writing skills. *Modern Language Journal, 75*, 305–313.

Kiefer, K., & Neufeld, J. (2002). Making the most of response: Reconciling coaching and evaluating roles for teachers across the curriculum. *Academic Writing, 3*. Retrieved from http://wac.colostate.edu/aw/articles/kiefer_neufeld_2002.pdf

Kim, L. (2004). Online technologies for teaching writing: Students react to teacher response in voice and written modalities. *Research in the Teaching of English, 38*(3), 304-337.

Knutsson, O., Pargman, T. C., Eklundh, K. S., & Westlund, S. (in press). Designing and developing a language environment for second language writers. *Computers & Education*.

Kouritzin, S. G., & Vizard, C. (1999). Feedback on feedback: Preservice ESL teachers respond to evaluation practices. *TESL Canada Journal, 17*(1). 16-39.

Kubota, M. (2001). Error correction strategies used by learners of Japanese when revising a writing task. *System, 29*, 467–480.

Kumari Dheram, P. (1995). Feedback as a two-bullock cart: A case study of teaching writing. *ELT Journal, 49*(2), 160–168.

Lai, P. C. (1986). The revision processes of first-year students at the National University of Singapore. *RELC Journal, 17*, 71–84.

*Lalande, J. F. (1982). Reducing composition errors: An experiment. *Modern Language Journal, 66*, 140-149.

Lane, T., & Potter, B. (1998). *Teaching collaborative feedback strategies in intermediate writing*. Retrieved from ERIC database. (ED427530)

*Lee, I. (1997). ESL learners' performance in error correction in writing: Some implications for teaching. *System, 25*, 465–477

Lee, I. (2003). L2 writing teachers' perspectives, practices and problems regarding error feedback. *Assessing Writing, 8*, 216–237.

Lee, I. (2004). Error correction in L2 secondary writing classrooms: The case of Hong Kong. *Journal of Second Language Writing, 13*, 285–312.

Lee, I. (2005). Error correction in the L2 writing classroom: What do students think? *TESL Canada Journal, 22*(2), 1–16.

Leki, I. (1990). Coaching from the margins: Issues in written response. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 57–68). Cambridge, UK: Cambridge University Press.

Leki, I. (1991). The preferences of ESL students for error correction in college level writing classes. *Foreign Language Annals, 23*, 203–218.

Leppänen, S., & Kalaja, P. (1995). Experimenting with computer conferencing in English for Academic Purposes. *ELT Journal, 49*(1), 26-36.

Levine, A., Oded, B., Connor, U., & Asons, I. (2002). Variation in EFL-ESL peer response. *TESL-EJ, 6*(3), 1–18.

Lin, D. (2001). *How mind works to revise compositions*. Retrieved from ERIC database. (ED 472654)

Lindgren, E. (2004). The uptake of peer-based intervention in the writing classroom. In G. Rijlaarsdam (Series Ed.) & G. Rijlaarsdam, H. Van den Bergh, & M. Couzijn (Vol. Eds.), *Studies in writing, 14*, (pp. 259–274). Norwell, MA: SpringerLiou, H.-C. (1992). An automatic text-analysis project for EFL writing revision. *System, 20*, 481–492.

Liu, J., & Sandler, R. W. (2003). The effect and affect of peer review in electronic versus traditional modes of L2 writing. *Journal of English for Academic Purposes, 2*, 193–227.

Lockhart, C., & Ng, P. (1995). Analyzing talk in ESL peer response groups: Stances, functions, and content. *Language Learning, 45*, 605–655.

MacDonald, R. (1991). *Developmental students' processing of teacher feedback in composition instruction*. Retrieved from ERIC database. (ED354965)

Mangelsdorf, K., & Schlumberger, A. (1992). ESL student response stances in a peer-review task. *Journal of Second Language Writing, 1*, 235–254.

Mathison Fife, J., & O'Neill, P. (2001). Moving beyond the written comment: Narrowing the gap between response practice and research. *College Composition and Communication, 53*, 300–320.

*Matsumura, L. C., Patthey-Chavez, G. G., Valdes, R., & Garnier, H. (2002). Teacher feedback, writing assignment quality, and third-grade students' revision in lower- and higher-achieving urban schools. *Elementary School Journal, 103*(1), 3–22.

Matsumura, S., & Hann, G. (2004). Computer anxiety and students' preferred feedback methods in EFL writing. *Modern Language Journal, 88,* 403–415.

*McCutchen, D., Francis, M., & Kerr, S. (1997). Revising for meaning: Effects of knowledge and strategy. *Journal of Educational Psychology, 89,* 667–676.

McCutchen, D., Hull, G. A., & Smith, W. L. (1987). Editing strategies and error correction in basic writing. *Written Communication, 4,* 139–152.

McGarrell, H., & Verbeem, J. (2007). Motivating revision of drafts through formative feedback. *ELT Journal 61*(3), 228–236.

*McGroarty, M. E., & Zhu, W. (1997). Triangulation in classroom research: A study of peer revision. *Language Learning, 47,* 1–43.

Medcalf, J., Glynn, T., & Moore, D. (2004). Peer tutoring in writing: A school systems approach. *Educational Psychology in Practice, 20*(2), 157–178.

Melby-Mauer, J. (2003). Using e-mail assignments and online correction in ESL instruction. *TESOL Journal, 12*(2), 37–38.

Mellen, C., & Sommers, J. (2003). Audiotaped response and the two-year-campus writing classroom: The two-sided desk, the "guy with the ax," and the chirping birds. *Teaching English in the Two Year College, 31*(1), 25–39.

Mendonca, C. O., & Johnson, K. E. (1994). Peer review negotiations: Revision activities in ESL writing instruction. *TESOL Quarterly, 28,* 745–768.

Miao, Y., Badger, R., & Zhen, Y. (2006). A comparative study of peer and teacher feedback in a Chinese EFL writing class. *Journal of Second Language Writing, 15,* 179–200.

Michaud, L. N., McCoy, K. F., & Pennington, C. A. (2000, November). An intelligent tutoring system for deaf learners of written English. Paper presented at Assets'00 Conference, Arlington, VA.

Min, H.-T. (2005). Training students to become successful peer reviewers. *System, 33,* 293–308.

Min, H.-T. (2006). The effects of trained peer review on EFL students' revision types and writing quality. *Journal of Second Language Writing, 15,* 118–141

Morris, L. A. (1998). Differences in men's and women's ESL writing at the junior college level: Consequence for research on feedback. *Canadian Modern Language Review, 55,* 219–238.

Muncie, J. (2000). Using written teacher feedback in EFL composition classes. *ELT Journal, 54*, 47–53.

Murphy, S. (2000). A sociocultural perspective on teacher response: Is there a student in the room? *Assessing Writing, 7*, 79–90.

Myers, S. (1997). Teaching writing as a process and teaching sentence-level syntax: Reformulation as ESL composition feedback. *TESL-EJ, 2*(4), 1–16.

Myles, J. (2002). Second language writing and research: The writing process and error analysis in student texts. *TESL-EJ, 6*(2), 1–20.

Nagata, N. (1993). Intelligent computer feedback for second language instruction. *Modern Language Journal, 77*, 330–339.

Nelson, G. L., & Carson, J. G. (1998). ESL students' perceptions of effectiveness in peer response groups. *Journal of Second Language Writing, 7*, 113–131.

Nelson, G. L., & Murphy, J. M. (1992). An L2 writing group: Task and social dimensions. *Journal of Second Language Writing, 1*, 171–193.

Nelson, G. L., & Murphy, J. M. (1993). Peer response groups: Do L2 writers use peer comments in revising their drafts? *TESOL Quarterly, 27*, 135–141.

Norton, L., & Norton, J. (2001). *Essay feedback: How can it help students improve their academic writing?* Retrieved from ERIC database. (ED454530)

*Olson, M., & Raffeld, P. (1987). The effects of written comments on the quality of student compositions and the learning of content. *Reading Psychology: An International Quarterly, 8*, 273–293.

Patthey-Chavez, G. G., & Ferris, D. R. (1997). Writing conferences and the weaving of multi-voiced texts in college composition. *Research in the Teaching of English, 31*(1), 51–90.

Paulson, E. J., Alexander, J., & Armstrong, S. (2007). Peer review re-viewed: Investigating the juxtaposition of composition students' eye movements and peer-review processes. *Research in the Teaching of English, 41*(3), 304-335.Paulus, T. M. (1999). The effect of peer and teacher feedback on student writing. *Journal of Second Language Writing, 8*, 265–289.

Peckham, I. (1996). If it ain't broke, why fix it?: Disruptive and constructive computer-mediated response group practices. *Computers and Composition, 13*, 327–339.

Pennington, M. C. (1993). A critical examination of word processing effects in relation to L2 writers. *Journal of Second Language Writing, 2*, 227–255.

Perpignan, H. (2003). Exploring the written feedback dialogue: A research, learning and teaching practice. *Language Teaching Research, 7*, 259–278.

Peterson, S., Childs, R., & Kennedy, K. (2004). Written feedback and scoring of sixth-grade girls' and boys' narrative and persuasive writing. *Assessing Writing, 9*, 160–180.

Peterson, S. S., & Kennedy, K. (2006). Sixth-grade teachers' written comments on student writing: Genre and gender influences. *Written Communication, 23*, 36–62.

Phelps, L. W. (2000). Cyrano's nose: Variations on the theme of response. *Assessing Writing, 7*, 91–110.

Phinney, M., & Khouri, S. (1993). Computers, revision, and ESL writers: The role of experience. *Journal of Second Language Writing, 2*, 257–277.

*Polio, D., Fleck, C., & Leder, N. (1998). "If I only had more time:" ESL learners' changes in linguistic accuracy on essay revisions. *Journal of Second Language Writing, 7,* 43–68.

Porte, G. (1997). The etiology of poor second language writing: The influence of perceived teacher preferences on second language revision strategies. *Journal of Second Language Writing*, 6, 61–78.

Porte, G. (2001). Losing sight of errors: The effects of typographical conditions on error salience in L2 proof-reading. *System, 29*, 137–148.

Porto, M. (2001). Cooperative writing response groups and self-evaluation. *ELT Journal, 55*, 38–46.

Powers, J. K., & Nelson, J. V. (1995). L2 writers and the writing center: A national survey of writing center conferencing at graduate institutions. *Journal of Second Language Writing, 4*, 113–138.

Pujola, J-T. (2001). Did CALL feedback feed back? Researching learners' use of feedback. *ReCALL, 13*, 79–98.

Qi, D. S., & Lapkin, S. (2001). Exploring the role of noticing in a three-stage second language writing task. *Journal of Second Language Writing, 10*, 277–303.

Radecki, P. M., & Swales, J. M. (1988). ESL student reaction to written comments on their written work. *System, 16*, 355–365.

Reed, M., & Burton, J. (1986). Effective and ineffective evaluation of essays: Perceptions of college freshmen. *Journal of Teaching Writing, 4*(2), 270–283. Reid, J. (1994). Responding to ESL students' texts: The myths of appropriation. *TESOL Quarterly, 28,* 273–292.

Richardson, S. (2000). Students' conditioned response to teachers' response. Portfolio proponents, take note! *Assessing Writing, 7*, 117–141.

Rieber, L. (2006). Using peer review to improve student writing in business courses. *Journal of Education for Business*, 82(6), 322–326.

Riedel, E., Dexter, S., Scharber, C., & Doering, A. (2006). Experimental evidence on the effectiveness of automated essay scoring in teacher educated cases. *Journal of Educational Computing Research, 35*(3), 267–287.

Rilling, S. (2005). The development of an ESL OWL, or learning how to tutor writing online. *Computers and Composition, 22*, 357–374.

Rinnert, C., & Kobayashi, H. (2001). Differing perceptions of EFL writing among readers in Japan. *Modern Language Journal, 85*, 189–209.

Robb, T., Ross, S., & Shortreed, I. (1986). Salience of feedback on error and its effect on EFL writing quality. *TESOL Quarterly, 20*, 83–91.

Rock, J. L. (2007). *The impact of short-term use of Criterion on writing skills in ninth grade*. (ETS Research Report RR-07-07). Princeton, NJ: ETS.

Rollinson, P. (2005). Using peer feedback in the ESL writing class. *ELT Journal, 59*, 23–30.

Rothschild, D., & Klingenberg, F. (1990). Self and peer evaluation of writing in the interactive ESL classroom: An exploratory study. *TESL Canada Journal, 8*(1), 52–65

Rouiller, Y. (2004). Metacognitive regulations, peer interactions and revision of narratives by sixth-graders. In G. Rijlaarsdam (Series Ed.) & G. Rijlaarsdam, H. Van den Bergh, & M. Couzijn (Vol. Eds.), *Studies in Writing, 14* (pp. 77–89). Norwell, MA: Springer.

Sachs, R., & Polio, C. (2007). Learners' uses of two types of written feedback on a L2 writing revision task. *Studies in Second Language Acquisition, 29*, 67–100.

Saito, H. (1994). Teachers' practices and students' preferences for feedback on second language writing: A case study of adult ESL learners. *TESL Canada Journal, 11*(2), 46–70.

Schacter, J. (1991). Corrective feedback in historical perspective. *Second Language Research, 7*(2), 89–102.

Schultz, J. M. (2000). Computers and collaborative writing in the foreign language classroom. In M. Warschauer & R. Kern (Eds.), *Network-based language learning: Concepts and practice* (pp. 121–150). Cambridge, UK: Cambridge University Press.

Schunk, D., & Swartz, C. (1993a). *Goals and progress feedback: Effects on self-efficacy and writing achievement*. Retrieved from ERIC database. (ED359216)

Schunk, D., & Swartz, C. (1993b). *Writing strategy instruction with gifted students: effects of goal and feedback on self-efficacy and skills*. Retrieved from ERIC database. (EJ465444)

Semke, H. D. (1984). The effects of the red pen. *Foreign Language Annals, 17*, 195–202.

Sengupta, S. (1998). Peer evaluation: "I am not the teacher." *ELT Journal, 52*, 19–28.

*Sengupta, S. (2000). An investigation into the effects of revision strategy instruction on L2 secondary school learners. *System, 28*, 97–113.

Severino, C. (1993). The sociopolitical implications of response to second language and second dialect writing. *Journal of Second Language Studies, 2*, 181–201.

Sheppard, K. (1992). Two feedback types: Do they make a difference? *RELC Journal, 23*, 103–110.

Shin, S. J. (2003). The reflective L2 writing teacher. *ELT Journal, 57*, 3–10.

Simpson, J. M. (2006). Feedback on writing: Changing EFL students' attitudes. *TESL Canada Journal, 24*(1), 96–112.

Sipple, S. (2007). Ideas in practice: Developmental writers' attitudes toward audio and written feedback. *Journal of Developmental Education, 30*(3), 22–31.

Smith, S. (1997). The genre of the end comment: Conventions in teacher responses to student writing. *College Composition and Communication, 48*, 249–268.

Smith Taylor, S., & Patton, M. D. (2006). Ten engineers reading: Disjunctions between preference and practice in civil engineering faculty responses. *Journal of Technical Writing and Communication, 36*, 253–271.

Sommers, N. (1980). Revision strategies of student writers and experienced adult writers. *College Composition and Communication, 31*, 378–388.

Sommers, N. (1982). Responding to student writing. *College Composition and Communication, 33*, 148–156.

Sommers, N. (2006). Across the drafts. *College Composition and Communication, 58*, 248–257.

Song, M. (1998). *Experimental study of the effect of controlled vs. free writing and different feedback types on writing quality and writing apprehension of EFL college students*. Retrieved from ERIC database. (ED423703)

Sperling, M., & Freedman, S. W. (1987). A good girl writes like a good girl: Written responses to student writing. *Written Communication, 9*(9), 337–355.

Stanley, J. (1992). Coaching student writers to be effective peer evaluators. *Journal of Second Language Writing, 1*, 217–233.

Stern, L.A., & Solomon, A. (2006). Effective faculty feedback: The road less traveled. *Assessing Writing, 11*, 22–41.

Storch, N., & Tapper, J. (1996). Patterns of NNS student annotations when identifying areas of concern in their writing. System, 24, 323–336.

Storch, N., & Tapper, J. (2000). The focus of teacher and student concerns in discipline-specific writing by university students. *Higher Education Research & Development, 19*, 337–355.

Straub, R. (1996). The concept of control in teacher response: Defining the varieties of "directive" and "facilitative" commentary. *College Composition and Communication, 47*(2), 223–251.

Straub, R. (1997). Students' reactions to teacher comments: An exploratory study. *Research in the Teaching of English, 31*(1), 91–119.

Straub, R. (2000). The student, the text, and the classroom context: A case study of teacher response. *Assessing Writing, 7*, 23–55.

Strenski, E., O'Dwyer Feagin, C., & Singer, J. A. (2005). Email small group peer review revisited. *Computers and Composition, 22*, 191–208.

Sugita, Y. (2006). The impact of teacher' comment types on students' revision. *ELT Journal, 60*(1), 34–41.

Sullivan, N., & Pratt, E. (1996). A comparative study of two ESL writing environments: a computer-assisted classroom and a traditional oral classroom. *System, 29*(4), 491–501.

Sullivan, K., & Lindgren, E. (2002). Self-assessment in autonomous computer-aided second language writing. *ELT Journal, 56*, 258–266.

Sweeney, M. R. (1999). Relating revision skills to teacher commentary. *Teaching English in the Two Year College, 27*, 213–218.

Tang, G. M., & Tithecott, J. (1999). Peer response and ESL writing. *TESL Canada Journal, 16*(2), 20–38.

Tang, R. (2000). Do we allow what we encourage? How students are positioned by teacher feedback. *Australian Journal of Language and Literacy, 23*, 157–168.

Thonus, T. (2001). Triangulation in the writing center: Tutor, tutee, and instructor perceptions of the tutor's role. *Writing Center Journal, 22*(1), 59–82.

Thonus, T. (2002). Tutor and student assessments of academic writing tutorials: What is "success"? *Assessing Writing, 8*, 110–134.

Todd, R. W., Mills, N., Palard, C., & Khamcharoen, P. (2001). Giving feedback on journals. *ELT Journal, 55*, 354–359.

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research, 68*, 249–276.

Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning, 46*, 327–369.

Truscott, J. (1999). The case for "The case against grammar correction in L2 writing classes": A response to Ferris. *Journal of Second Language Writing, 8*, 111–122.

Truscott, J. (2004). Evidence and conjecture on the effects of correction: A response to Chandler. *Journal of Second Language Writing, 13*, 337–343.

Tsui, A. B. M., & Ng, M. (2000). Do secondary L2 writers benefit from peer comments? *Journal of Second Language Writing, 9*, 147–170. Tuzi, F. (2004). The impact of e-feedback on the revisions of L2 writers in an academic writing course. *Computers and Composition, 21*, 217–235.

Van Gelderen, A. (1997). Elementary students' skills in revising: Integrating quantitative and qualitative analysis. *Written Communication, 14*(3), 360–397.

Villamil, O. S., & De Guerrero, M. C. M. (1998). Assessing the impact of peer revision on L2 writing. *Applied Linguistics, 19*, 491–514.

Wallace, D. L., & Hayes, J. R. (1991). Redefining revision for freshmen. *Research in the Teaching of English, 25*, 54–66.

Wallace, D. L., Hayes, J. R., Hatch, J. A., Miller, W., Moser, G., & Silk, C. M. (1996). Better revision in eight minutes? Prompting first-year college writers to revise globally. *Journal of Education Psychology, 88*, 682–688.

Warden, C. A., (2000). EFL business writing behaviors in differing feedback environments. *Language Learning, 50*, 573–616.

Warden, C. A., & Chen, J. (1995). Improving feedback while decreasing teacher burden in R.O.C. ESL business English writing classes. In P. Bruthiaux, T. Boswood, & B. Du-Babcock (Eds.), *Explorations in English for professional communications* (pp. 125–137). Hong Kong, People's Republic of China: City University of Hong Kong.

Waring, H. Z. (2005). Peer tutoring in a graduate writing centre: Identity, expertise, and advice resisting. *Applied Linguistics, 26*, 141–168.

Weigle, S. C., & Nelson, G. L. (2004). Novice tutors and their ESL tutees: Three case studies of tutor roles and perceptions of tutorial success. *Journal of Second Language Writing, 13*, 203–225.

Weltig, M. S. (2004). Effects of language errors and importance attributed to language on language and rhetorical-level essay scoring. *Spaan Fellow Working Papers in Second or Foreign Language Assessment, 2*, 53–81.

Whetherbee Phelps, L. (2000). Cyrano's nose: Variations on the theme of response. *Assessing Writing, 7*, 91–110.

Whitington, V., Glover, A., & Harley, F. (2004). Preservice early childhood students' perceptions of written feedback on their essays. *Early Child Development and Care, 174*(4), 321–337.

Wible, D., Kuo, C.-H., Chien, F.-Y., Liu, A., & Tsao, N.-L. (2001). A web-based EFL writing environment: Integrating information for learners, teachers, and researchers. *Computers & Education, 37*, 297–315.

Williams, J. (2002). Undergraduate second language writers in the writing center. *Journal of Basic Writing, 21*(2), 73–91.

Williams, J. (2004). Tutoring and revision: Second language writers in the writing center. *Journal of Second Language Writing, 13*, 173–201.

Wiltse, E. (2001). *The effects of motivation and anxiety on students' use of instructor comments*. Retrieved from ERIC database. (ED458630)

Wiltse, E. (2002). Correlates of college students' use of instructors' comments. *Journalism and Mass Communication Educator, 57*(2),126–138.

*Xiang, W. (2004). Encouraging self-monitoring in writing by Chinese students. *ELT Journal, 58*, 238–246.

Yagelski, R. (1995). The role of classroom context in the revision strategies of student writers. *Research in the Teaching of English, 29*, 216–338.

Yang, M., Badger, R., & Yu, Z. (2006). A comparative study of peer and teacher feedback in a Chinese EFL writing class. *Journal of Second Language Writing, 15*, 179–200.

Yao, Y.-C., & Warden, C. (1996). Process writing and computer correction: Happy wedding or shotgun marriage? *CALL Electronic Journal, 1*(1). Retrieved from http://www.tell.is.ritsumei.ac.jp/callej/1-1/Warden1.html

Yates, R., & Kenkel, J. (2002). Responding to sentence-level errors in writing. *Journal of Second Language Writing, 11*, 29–47.

Zamel, V. (1985). Responding to student writing. *TESOL Quarterly, 19*, 79–101.

Zellermayer, M. (1989). The study of teachers' written feedback to students' writing: Changes in theoretical considerations and the expansion of research contexts. *Instructional Science 18*, 145–165.

Zhang, S. (1995). Reexamining the affective advantage of peer feedback in the ESL writing class. *Journal of Second Language Writing, 4*, 209–222.

Zhang, S. (1999). Thoughts on some recent evidence concerning the affective advantage of peer feedback. *Journal of Second Language Writing, 8*, 321–326.

Zhu, W. (1995). Effects of training for peer response on students' comments and interaction. *Written Communication, 12*, 492–528.

Zhu, W. (2001). Interaction and feedback in mixed peer response groups. *Journal of Second Language Writing, 10*, 251–276.

Zhu, W. (2004). Faculty views on the importance of writing, the nature of academic writing, and teaching and responding to writing in the disciplines. *Journal of Second Language Writing, 13*, 29–48.

Zimmerman, B. J., & Kitsantas, A. (1999). Acquiring writing revision skill: Shifting from process to outcome self-regulatory skills. *Journal of Education Psychology, 91*, 241–250.

*Zimmerman, B. J., & Kitsantas, A. (2002). Acquiring writing revision and self-regulatory skill through observation and emulation. *Journal of Educational Psychology, 94*, 660–668.

Ziv, N. (1984). The effect of teacher comments on the writing of four college freshmen. In R. Beach & L. Bridwell, (Eds.), *New directions in composition research* (pp. 362–380). New York, NY: Guilford.

Appendix B

Summary of All Individual Effect Sizes Included in the Quantitative Meta-Analysis

Study	Design	Description of groups	Sample size (*n*)	Type of feedback	Feedback focus	Outcome measure	Effect size
Ashwell (2000)	PP	Content then form	12; Group 1	LO	C, F	WP: accuracy: number of errors/ number of words	1.7
	PP	Form then content	13; Group 2	LO	F, C	WP: accuracy: number of errors/ number of words	-.75
	PP	F&C then F&C	13; Group 3	LO	F, C	WP: accuracy: number of errors/ number of words	1.28
	PP	Content then form	12; Group 1	LO	C, F	WP: content scores (rating 1-20)	.17
	PP	Form then content	13; Group 2	LO	F, C	WP: content scores (rating 1-20)	-.08
	PP	F&C then F&C	13; Group 3	LO	F, C	WP: content scores (rating 1-20)	-.34
Berg (1999)	TC	Trained vs. untrained, level 3	24	Revision activities	Revision	RV: meaning change	1.51
	TC	Trained vs. untrained, level 4	22	Revision activities	Revision	RV: meaning change	2.25
	TC	Trained vs. untrained, both levels	46	Revision activities	Revision	RV: meaning change	1.90
Bitchener, Young, & Cameron (2005)	TC	Oral and written vs. no feedback, time 1	36; Group 1	LO + DC	F	WP: % of correct uses (prepositions)	.21
	TC	Oral and written vs. no feedback, time 2	36; Group 1	LO + DC	F	WP: % of correct uses (prepositions)	.03
	TC	Oral and written vs. no feedback, time 3	36; Group 1	LO + DC	F	WP: % of correct uses (prepositions)	-.30

Study	Design	Description of groups	Sample size (*n*)	Type of feedback	Feedback focus	Outcome measure	Effect size
	TC	Oral and written vs. no feedback, time 4	36; Group 1	LO + DC	F	WP: % of correct uses (prepositions)	.88
	TC	Oral and written vs. no feedback, time 1	36; Group 1	LO + DC	F	WP: % of correct uses (past simple)	1.17
	TC	Oral and written vs. no feedback, time 2	36; Group 1	LO + DC	F	WP: % of correct uses (past simple)	.24
	TC	Oral and written vs. no feedback, time 3	36; Group 1	LO + DC	F	WP: % of correct uses (past simple)	-.25
	TC	Oral and written vs. no feedback, time 4	36; Group 1	LO + DC	F	WP: % of correct uses (past simple)	.77
	TC	Oral and written vs. no feedback, time 1	36; Group 1	LO + DC	F	WP: % of correct uses (definite article)	.23
	TC	Oral and written vs. no feedback, time 2	36; Group 1	LO + DC	F	WP: % of correct uses (definite article)	.83
	TC	Oral and written vs. no feedback, time 3	36; Group 1	LO + DC	F	WP: % of correct uses (definite article)	-.02
	TC	Oral and written vs. no feedback, time 4	36; Group 1	LO + DC	F	WP: % of correct uses (definite article)	1.76
	TC	Written only vs. control, time 1	34; Group 2	LO + DC	F	WP: % of correct uses (prepositions)	.32
	TC	Written only vs. control, time 2	34; Group 2	LO + DC	F	WP: % of correct uses (prepositions)	.10
	TC	Written only vs. control, time 3	34; Group 2	LO + DC	F	WP: % of correct uses (prepositions)	.04
	TC	Written only vs. control, time 4	34; Group 2	LO + DC	F	WP: % of correct uses (prepositions)	-.17
	TC	Written only vs. control, time 1	34; Group 2	LO + DC	F	WP: % of correct uses (past simple)	.30

Study	Design	Description of groups	Sample size (*n*)	Type of feedback	Feedback focus	Outcome measure	Effect size
	TC	Written only vs. control, time 2	34; Group 2	LO + DC	F	WP: % of correct uses (past simple)	-.57
	TC	Written only vs. control, time 3	34; Group 2	LO + DC	F	WP: % of correct uses (past simple)	-.27
	TC	Written only vs. control, time 4	34; Group 2	LO + DC	F	WP: % of correct uses (past simple)	-.37
	TC	Written only vs. control, time 1	34; Group 2	LO + DC	F	WP: % of correct uses (definite article)	.47
	TC	Written only vs. control, time 2	34; Group 2	LO + DC	F	WP: % of correct uses (definite article)	.43
	TC	Written only vs. control, time 3	34; Group 2	LO + DC	F	WP: % of correct uses (definite article)	.11
	TC	Written only vs. control, time 4	34; Group 2	LO + DC	F	WP: % of correct uses (definite article)	.65
Cardelle & Corno (1981)	PP	Written, praise, pretest vs. posttest 1, course 1	12; Group 1	SE + SI	F, C	WP scores (grammar + vocabulary)	.41
	PP	Written, praise, pretest vs. posttest 2, course 1	12; Group 1	SE + SI	F, C	WP scores (grammar + vocabulary)	.72
	PP	Written, praise, pretest vs. posttest 3, course 1	12; Group 1	SE + SI	F, C	WP scores (grammar + vocabulary)	.82
	PP	Written, praise, pretest vs. posttest 1, course 2	7; Group 2	SE + SI	F, C	WP scores (grammar + vocabulary)	1.75
	PP	Written, praise, pretest vs. posttest 2, course 2	7; Group 2	SE + SI	F, C	WP scores (grammar + vocabulary)	1.81
	PP	Written, praise, pretest vs. posttest 3, course 2	7; Group 2	SE + SI	F, C	WP scores (grammar + vocabulary)	1.66
	PP	Written, criticism, pretest vs. posttest 1, course 1	8; Group 3	SE + ML	F, C	WP scores (grammar + vocabulary)	2.09

Study	Design	Description of groups	Sample size (*n*)	Type of feedback	Feedback focus	Outcome measure	Effect size
	PP	Written, criticism, pretest vs. posttest 2, course 1	8; Group 3	SE + ML	F, C	WP scores (grammar + vocabulary)	2.34
	PP	Written, criticism, pretest vs. posttest 3, course 1	8; Group 3	SE + ML	F, C	WP scores (grammar + vocabulary)	2.34
	PP	Written, criticism, pretest vs. posttest 1, course 2	6; Group 4	SE + ML	F, C	WP scores (grammar + vocabulary)	.65
	PP	Written, criticism, pretest vs. posttest 2, course 2	6; Group 4	SE + ML	F, C	WP scores (grammar + vocabulary)	1.92
	PP	Written, criticism, pretest vs. posttest 3, course 2	6; Group 4	SE + ML	F, C	WP scores (grammar + vocabulary)	.67
	PP	Combination (praise + criticism), pretest vs. posttest1, course 1	12; Group 5	SE, SI, ML	F, C	WP scores (grammar + vocabulary)	3.41
	PP	Combination (praise + criticism), pretest vs. posttest2, course 1	12; Group 5	SE, SI, ML	F, C	WP scores (grammar + vocabulary)	4.22
	PP	Combination (praise + criticism), pretest vs. posttest3, course 1	12; Group 5	SE, SI, ML	F, C	WP scores (grammar + vocabulary)	4.44
	PP	Combination (praise + criticism), pretest vs. posttest1, course 2	7; Group 6	SE, SI, ML	F, C	WP scores (grammar + vocabulary)	.62
	PP	Combination (praise + criticism), pretest vs. posttest2, course 2	7; Group 6	SE, SI, ML	F, C	WP scores (grammar + vocabulary)	.75
	PP	Combination (praise + criticism), pretest vs. posttest3, course 2	7; Group 6	SE, SI, ML	F, C	WP scores (grammar + vocabulary)	.64

Study	Design	Description of groups	Sample size (*n*)	Type of feedback	Feedback focus	Outcome measure	Effect size
	TC	Praise vs. no feedback, avg. posttest, course 1	22	SE, SI	F, C	WP scores (grammar + vocabulary)	.76
	TC	Praise vs. no feedback, avg. posttest, course 2	13	SE, SI	F, C	WP scores (grammar + vocabulary)	1.07
	TC	Criticism vs. no feedback, avg. posttest, course 1	18	SE, ML	F, C	WP scores (grammar + vocabulary)	1.13
	TC	Criticism vs. no feedback, avg. posttest, course 2	12	SE, ML	F, C	WP scores (grammar + vocabulary)	.88
	TC	Combination vs. no feedback, avg. posttest, course 1	22	SE, SI, ML	F, C	WP scores (grammar + vocabulary)	2.21
	TC	Combination vs. no feedback, avg. posttest, course 2	13	SE, SI, ML	F, C	WP scores (grammar + vocabulary)	1.07
Chandler (2003), Study 1	PP	Direct correction, ch. 1-5	15; Group 1	DC, CM	F, C	Accuracy: number of errors per 100 words	1.04
	PP	Direct correction, ch. 1-5	14; Group 2	DC, CM	F, C	Fluency: minutes per 100 words	.82
	TC	Direct correction vs. control, ch 1	31; Group 3	DC, CM	F, C	Accuracy: number of errors per 100 words (did not decrease)	-.49
	TC	Direct correction vs. control, ch 5	31; Group 3	DC, CM	F, C	Accuracy: number of errors per 100 words	.52
	TC	Direct correction vs. control, ch 1	26; Group 4	DC, CM	F, C	Fluency: minutes per 100 words	.02
	TC	Direct correction vs. control, ch 5	26; Group 4	DC, CM	F, C	Fluency: minutes per 100 words (did not increase)	-.04

Study	Design	Description of groups	Sample size (*n*)	Type of feedback	Feedback focus	Outcome measure	Effect size
Chandler (2003), Study 2	PP	Revision group, ch 1-5	65	DC, CM, LO	F, C	Accuracy: number of errors per 100 words	.43
	PP	Revision group, ch 1-5	30	DC, CM, LO	F, C	Fluency: mean time to write 100 words	1.77
	TC	Correction vs. description (treated as control)	61	DC	F, C	Accuracy: number of errors per 100 words	1.02
	TC	Underlining and description vs. description	54	LO+CM	F, C	Accuracy: number of errors per 100 words	.42
	TC	Underlining vs. description	57	LO	F, C	Accuracy: number of errors per 100 words	.05
Davis & Fulton (1997)	PP	Feedback during composing process	20	SE	F, C	Writing quality: rating scales	1.86
	PP	Feedback after composing process	20	SE	F, C	Writing quality: rating scales	2.47
Davis & Kelley (1999)	PP	Feedback during composing process	45	SE	F, C	Writing quality: rating scales	1.80
	PP	Feedback after composing process	43	SE	F, C	Writing quality: rating scales	1.54
Fathman & Whalley (1990)	PP	Grammar feedback	14	LO	F	Writing quality: grammar score (number of errors)	2.08
	PP	Grammar feedback	14	LO	F	Writing quality: content rating	1.04
	PP	Content feedback	22	CM	C	Writing quality: grammar score (number of errors)	.09
	PP	Content feedback	22	CM	C	Writing quality: content rating	1.22
	PP	Grammar + content	22	LO, CM	F, C	Writing quality: grammar score (number of errors)	2.15

Study	Design	Description of groups	Sample size (*n*)	Type of feedback	Feedback focus	Outcome measure	Effect size
	PP	Grammar + content	22	LO, CM	F, C	Writing quality: content rating	1.02
Ferris (2006)	PP	Grammar feedback, essay 1-4	55	EC	F	Writing quality: total errors	.36
	PP	Grammar feedback, essay 1-4	55	EC	F	Writing quality: verb errors	.50
	PP	Grammar feedback, essay 1-4	55	EC	F	Writing quality: noun errors	.20
	PP	Grammar feedback, essay 1-4	55	EC	F	Writing quality: article errors (did not identify more)	-.10
	PP	Grammar feedback, essay 1-4	55	EC	F	Writing quality: lexical errors	.26
	PP	Grammar feedback, essay 1-4	55	EC	F	Writing quality: sentence errors	-.01
Ferris & Roberts (2001)	TC	Codes vs. no feedback, verbs	42	LO + EC	F	Writing quality: errors marked	.81
	TC	Codes vs. no feedback, nouns	42	LO + EC	F	Writing quality: errors marked	.69
	TC	Codes vs. no feedback, articles	42	LO + EC	F	Writing quality: errors marked	.06
	TC	Codes vs. no feedback, word choice	42	LO + EC	F	Writing quality: errors marked (did not identify more)	-.47
	TC	Codes vs. no feedback, sentence structure	42	LO + EC	F	Writing quality: errors marked	.10
	TC	Codes vs. no feedback, total	42	LO + EC	F	Writing quality: errors marked	.64
	TC	No codes vs. no feedback, verbs	39	LO	F	Writing quality: errors marked	.67

Study	Design	Description of groups	Sample size (*n*)	Type of feedback	Feedback focus	Outcome measure	Effect size
	TC	No codes vs. no feedback, nouns	39	LO	F	Writing quality: errors marked	.77
	TC	No codes vs. no feedback, articles	39	LO	F	Writing quality: errors marked	.16
	TC	No codes vs. no feedback, word choice	39	LO	F	Writing quality: errors marked	-.49
	TC	No codes vs. no feedback, sentence structure	39	LO	F	Writing quality: errors marked	.09
	TC	No codes vs. no feedback, total	39	LO	F	Writing quality: errors marked	.46
Hillocks (1982)	PP	Prewriting + revision	75	CM	F, C	Writing quality: rating scale	.61
	PP	Prewriting, no revision	72	-	-	Writing quality: rating scale	.82
	PP	Assignment + revision	67	CM	F, C	Writing quality: rating scale	.81
Kamimura (2006)	PP	Oral peer feedback; high proficiency group	12	SI	-	Writing quality: rating scale	1.96
	PP	Oral peer feedback; low proficiency group	12	SI	-	Writing quality: rating scale	2.33
Lalande (1982)	TC	Error-correction group vs. control	60	EC	F	Writing quality: number of errors	.46
	PP	Error-correction group	30	EC	F	Writing quality: number of non-lexical errors	.29
Lee (1997)	TC	Marked vs. unmarked (control)	99	LO	F, C	Error correction score	2.28
	TC	Slightly marked vs. unmarked	99	EX	F, C	Error correction score	2.24
Matsumura, Patthey-Chavez,	PP	Revision, lower achieving group	43	DC, CM	F	Writing quality: rating scale, content	.10

Study	Design	Description of groups	Sample size (*n*)	Type of feedback	Feedback focus	Outcome measure	Effect size
Valdes, & Garnier (2002)	PP	Revision, lower achieving group	43	DC, CM	F	Writing quality: rating scale, organization	.32
	PP	Revision, lower achieving group	43	DC, CM	F	Writing quality: rating scale, writing conventions	.85
	PP	Revision, higher achieving group	44	DC, CM	F	Writing quality: rating scale, content	.10
	PP	Revision, higher achieving group	44	DC, CM	F	Writing quality: rating scale, organization	.10
	PP	Revision, higher achieving group	44	DC, CM	F	Writing quality: rating scale, writing conventions	.31
McCutchen, Francis, & Kerr (1997), study 1	TC	Uncued vs. cued, familiar text, 7th graders	23	EX	F, C	Number of errors corrected, spelling (**uncued did not correct more**)	-.75
	TC	Uncued vs. cued, familiar text, 7th graders	23	EX	F, C	Number of errors corrected, meaning	.87
	TC	Uncued vs. cued, unfamiliar text, 7th graders	23	EX	F, C	Number of errors corrected, spelling	-.74
	TC	Uncued vs. cued, unfamiliar text, 7th graders	23	EX	F, C	Number of errors corrected, meaning	.08
	TC	Uncued vs. cued, familiar text, college students	14	EX	F, C	Number of errors corrected, spelling	-.37
	TC	Uncued vs. cued, familiar text, college students	14	EX	F, C	Number of errors corrected, meaning	-.37
	TC	Uncued vs. cued, unfamiliar text, college students	14	EX	F, C	Number of errors corrected, spelling	-.17

Study	Design	Description of groups	Sample size (*n*)	Type of feedback	Feedback focus	Outcome measure	Effect size
	TC	Uncued vs. cued, unfamiliar text, college students	14	EX	F, C	Number of errors corrected, meaning	-.55
McCutchen, Francis, & Kerr (1997), study 2	TC	Uncued vs. cued, familiar text, high ability	8	EX	F, C	Number of errors corrected, spelling	NA
	TC	Uncued vs. cued, familiar text, high ability	8	EX	F, C	Number of errors corrected, meaning	.99
	TC	Uncued vs. cued, unfamiliar text, high ability	8	EX	F, C	Number of errors corrected, spelling	-.99
	TC	Uncued vs. cued, unfamiliar text, high ability	8	EX	F, C	Number of errors corrected, meaning	NA
	TC	Uncued vs. cued, familiar text, middle ability	8	EX	F, C	Number of errors corrected, spelling	-.45
	TC	Uncued vs. cued, familiar text, middle ability	8	EX	F, C	Number of errors corrected, meaning	-.20
	TC	Uncued vs. cued, unfamiliar text, middle ability	8	EX	F, C	Number of errors corrected, spelling	NA
	TC	Uncued vs. cued, unfamiliar text, middle ability	8	EX	F, C	Number of errors corrected, meaning	-.50
	TC	Uncued vs. cued, familiar text, low ability	12	EX	F, C	Number of errors corrected, spelling	0
	TC	Uncued vs. cued, familiar text, low ability	12	EX	F, C	Number of errors corrected, meaning	.37

Study	Design	Description of groups	Sample size (*n*)	Type of feedback	Feedback focus	Outcome measure	Effect size
	TC	Uncued vs. cued, unfamiliar text, low ability	12	EX	F, C	Number of errors corrected, spelling	-.41
	TC	Uncued vs. cued, unfamiliar text, low ability	12	EX	F, C	Number of errors corrected, meaning	.82
McGroarty & Zhu (1997)	TC	Training in peer revision vs. no training	89	Revision	F, C	Holistic scores, round 1	.08
	TC	Training in peer revision vs. no training	89	Revision	F, C	Holistic scores, round 2	.06
	TC	Training in peer revision vs. no training	89	Revision	F, C	Holistic scores, portfolio grades	.25
Brakel Olson (1990)	PP	Revision + peer practice	23	Revision strategies	F	Writing quality: rating scale, total	.33
	PP	Revision	24	Revision strategies	F	Writing quality: rating scale, total	.14
	PP	Revision + peer practice	23	Revision strategies	F	Writing quality: rating scale, rhetorical quality	.23
	PP	Revision	24	Revision strategies	F	Writing quality: rating scale, rhetorical quality (did not improve)	-.03
	PP	Revision + peer practice	23	Revision strategies	F	Writing quality: rating scale, surface structure	.37
	PP	Revision	24	Revision strategies	F	Writing quality: rating scale, surface structure	.47

Study	Design	Description of groups	Sample size (*n*)	Type of feedback	Feedback focus	Outcome measure	Effect size
Olson and Raffeld (1987)	TC	Comments on content vs. no comments	16	CM	C	Writing quality: holistic	1.12
	TC	Comments on content vs. no comments	16	CM	C	Writing quality: content	.10
	TC	Comments on form vs. no comments	14	CM	F	Writing quality: holistic	.02
	TC	Comments on form vs. no comments	14	CM	F	Writing quality: content	-.90
Polio, Fleck, & Leder (1998)	PP	Error correction group, 30 minute	34	DC	F	Linguistic accuracy: error-free T-units (EFTs)	.35
	PP	Error correction group, 30 minute	34	DC	F	Linguistic accuracy: number of words in EFTs per total words	.32
	PP	Error correction group, 60 minute	34	DC	F	Linguistic accuracy: error-free T-units (EFTs)	.50
	PP	Error correction group, 60 minute	34	DC	F	Linguistic accuracy: number of words in EFTs per total words	.46
	TC	Error correction vs. control, 30 minute posttest	65	DC	F	Linguistic accuracy: error-free T-units (EFTs)	.10
	TC	Error correction vs. control, 30 minute posttest posttest	65	DC	F	Linguistic accuracy: number of words in EFTs per total words (did not perform better)	-.10
	TC	Error correction vs. control, 60 minute posttest	65	DC	F	Linguistic accuracy: error-free T-units (EFTs) (did not perform better)	-.12
	TC	Error correction vs. control, 60 minute posttest posttest	65	DC	F	Linguistic accuracy: number of words in EFTs per total words (did not perform better)	-.05

Study	Design	Description of groups	Sample size (*n*)	Type of feedback	Feedback focus	Outcome measure	Effect size
	PP	Revision group 1: error feedback + revision	35	DC, CM, revision	Not given	Holistic grading, no specific focus	1.19
	PP	Revision group 2: error feedback + revision	35	DC, CM, revision	Not given	Holistic grading, no specific focus	1.07
	PP	Traditional group: error feedback	30	DC, LO	Not given	Holistic grading, no specific focus	.37
Sengupta (2000)	TC	Revision group 1 vs. no feedback	35	Revision instruction		Holistic grading	.68
	TC	Revision group 2 vs. no feedback	35	Revision instruction		Holistic grading	.93
Xiang (2004)	TC	Experimental group (annotation) vs. control	58	SI, SE	F, C	Holistic scores: total	.30
	TC	Experimental group (annotation) vs. control	58	SI, SE	F, C	Content score	.30
	TC	Experimental group (annotation) vs. control	58	SI, SE	F, C	Holistic scores: organization	1.10
	TC	Experimental group (annotation) vs. control	58	SI, SE	F, C	Grammar score	.16
	TC	Experimental group (annotation) vs. control	58	SI, SE	F, C	Vocabulary score (did not perform better)	-.11
	TC	Experimental group (annotation) vs. control	58	SI, SE	F, C	Mechanics score	.13

Study	Design	Description of groups	Sample size (*n*)	Type of feedback	Feedback focus	Outcome measure	Effect size
Zimmerman & Kitsantas (2002)	TC	Social feedback vs. no social feedback, no model, writing skill posttest	24	SE	Not specified	Writing quality: rating scale	.73
	TC	Social feedback vs. no social feedback, mastery model, writing skill posttest	24	SE	Not specified	Writing quality: rating scale	.38
	TC	Social feedback vs. no social feedback, coping model, writing skill posttest	24	SE	Not specified	Writing quality: rating scale	.49
	TC	Social feedback vs. no social feedback, no model, self-efficacy posttest	24	SE	Not specified	Ability self-rating: rating scale	.61
	TC	Social feedback vs. no social feedback, mastery model, self-efficacy posttest	24	SE	Not specified	Ability self-rating: rating scale	-.02
	TC	Social feedback vs. no social feedback, coping model, self-efficacy posttest	24	SE	Not specified	Ability self-rating: rating scale	.98
	TC	Social feedback vs. no social feedback, no model, self-satisfaction posttest	24	SE	Not specified	Attitude self-rating: rating scale	.37

Study	Design	Description of groups	Sample size (*n*)	Type of feedback	Feedback focus	Outcome measure	Effect size
	TC	Social feedback vs. no social feedback, mastery model, self-satisfaction posttest	24	SE	Not specified	Attitude self-rating: rating scale	.06
	TC	Social feedback vs. no social feedback, coping model, self-satisfaction posttest	24	SE	Not specified	Attitude self-rating: rating scale	1.16

Note. See Table 1 for a list of all variables and a key to the variable codes.

Appendix C

Summary of the Final Effect Sizes Included in the Quantitative Meta-Analysis

Study	Design type	Language group	Level	Source of feedback	Mode of feedback	Type of feedback	Focus of feedback	Outcome type	Outcome focus	Effect size (or average effect size)
Ashwell (2000)	PP	E2		TE	WR	LO	C/F	WP	GR	1.7
	PP	E2		TE	WR	LO	C/F	WP	GR	-0.75
	PP	E2		TE	WR	LO	C/F	WP	GR	1.28
	PP	E2		TE	WR	LO	C/F	WP	C	0.17
	PP	E2		TE	WR	LO	C/F	WP	C	-0.08
	PP	E2		TE	WR	LO	C/F	WP	C	-0.34
Berg (1999)	TC	E2	H	O	WR	O	OTH	RV	C	1.51
	TC	E2	H	O	WR	O	OTH	RV	C	2.25
	TC	E2	H	O	WR	O	OTH	RV	C	1.9
Bitchener, Young, & Cameron (2005)	TC	E2	H	TE	WR	M	F	WP	GR	0.46[a]
	TC	E2	H	TE	WR	M	F	WP	GR	0.09[a]
Cardelle & Corno (1981)	PP	O2	L	TE	OR	CM	C/F	WP	GR	0.65[a]
	PP	O2	L	TE	OR	CM	C/F	WP	GR	1.74[a]
	PP	O2	L	TE	OR	CM	C/F	WP	GR	2.25[a]
	PP	O2	L	TE	OR	CM	C/F	WP	GR	1.08[a]
	PP	O2	L	TE	OR	CM	C/F	WP	GR	4.02[a]
	PP	O2	L	TE	OR	CM	C/F	WP	GR	0.67[a]
	TC	O2	L	TE	OR	CM	C/F	WP	GR	0.76
	TC	O2	L	TE	OR	CM	C/F	WP	GR	1.07
	TC	O2	L	TE	OR	CM	C/F	WP	GR	1.18
	TC	O2	L	TE	OR	CM	C/F	WP	GR	0.88
	TC	O2	L	TE	OR	CM	C/F	WP	GR	2.21
	TC	O2	L	TE	OR	CM	C/F	WP	GR	1.07

Study	Design type	Language group	Level	Source of feedback	Mode of feedback	Type of feedback	Focus of feedback	Outcome type	Outcome focus	Effect size (or average effect size)
Chandler (2003), study 1	PP	E2	H	TE	WR	M	C/F	WP	GR	1.04
	PP	E2	H	TE	WR	M	C/F	WP	H	0.82
	TC	E2	H	TE	WR	M	C/F	WP	GR	0.02[a]
	TC	E2	H	TE	WR	M	C/F	WP	H	-0.01[a]
Chandler (2003), study 2	PP	E2	H	TE	WR	M	C/F	WP	GR	0.43
	PP	E2	H	TE	WR	M	C/F	WP	H	1.77
	TC	E2	H	TE	WR	DC	C/F	WP	GR	1.02
	TC	E2	H	TE	WR	M	C/F	WP	GR	0.42
	TC	E2	H	TE	WR	LO	C/F	WP	GR	0.05
Davis & Fulton (1997)	PP	E1	H	TE	OR	CM	C/F	WP	H	1.86
	PP	E1	H	TE	OR	CM	C/F	WP	H	2.47
Davis & Mahoney (1999)	PP	E1	H	TE	OR	CM	C/F	WP	H	1.8
	PP	E1	H	TE	OR	CM	C/F	WP	H	1.54
Fathman & Whalley (1990)	PP	E2	L	TE	WR	LO	F	WP	GR	2.08
	PP	E2	L	TE	WR	LO	F	WP	C	1.04
	PP	E2	L	TE	WR	CM	C	WP	GR	0.09
	PP	E2	L	TE	WR	CM	C	WP	C	1.22
	PP	E2	L	TE	WR	M	C/F	WP	GR	2.15
	PP	E2	L	TE	WR	M	C/F	WP	C	1.02
Ferris (2006)	PP	E2		TE	WR	LO	F	WP	GR	0.2[a]
Ferris & Roberts (2001)	TC	E2		TE	WR	M	F	WP	GR	0.29[a]
	TC	E2		TE	WR	LO	F	WP	GR	0.28[a]
Hillocks (1982)	PP	E1		TE	WR	CM	C/F	WP	H	0.61
	PP	E1		TE	WR	CM	C/F	WP	H	0.81
Kamimura (2006)	PP	E2	H	O	OR	CM		WP	H	1.96
	PP	E2	L	O	OR	CM		WP	H	2.33
Lalande (1982)	TC	O2	H	TE	WR	LO	F	WP	H	0.46
	PP	O2	H	TE	WR	LO	F	WP	H	0.29

Study	Design type	Language group	Level	Source of feedback	Mode of feedback	Type of feedback	Focus of feedback	Outcome type	Outcome focus	Effect size (or average effect size)
Lee (1997)	TC	E2	L	O	WR	LO	C/F	WP	GR	2.28
	TC	E2	L	O	WR	M	C/F	WP	GR	2.24
Matsumura, Patthey-	PP	MX	L	TE	WR	M	F	WP	C	0.1
Chavez, Valdes, &	PP	MX	L	TE	WR	M	F	WP	H	0.32
Garnier (2002)	PP	MX	L	TE	WR	M	F	WP	GR	0.85
	PP	MX	H	TE	WR	M	F	WP	C	0.1
	PP	MX	H	TE	WR	M	F	WP	H	0.1
	PP	MX	H	TE	WR	M	F	WP	GR	0.31
McCutchen, Francis,	TC	E1	L	O	WR	CM	C/F	WP	SP	-0.75[a]
& Kerr (1997),	TC	E1	L	O	WR	CM	C/F	WP	C	0.48[a]
study 1	TC	E1	H	O	WR	CM	C/F	WP	SP	-0.27[a]
	TC	E1	H	O	WR	CM	C/F	WP	C	-0.46[a]
McCutchen, Francis,	TC	E1	H	O	WR	CM	C/F	WP	C	0.99[a]
& Kerr (1997),	TC	E1	H	O	WR	CM	C/F	WP	SP	-0.99[a]
study 2	TC	E1	M	O	WR	CM	C/F	WP	SP	-0.45[a]
	TC	E1	M	O	WR	CM	C/F	WP	C	-0.38[a]
	TC	E1	L	O	WR	CM	C/F	WP	SP	-0.2[a]
	TC	E1	L	O	WR	CM	C/F	WP	C	0.59[a]
McGroarty & Zhu (1997)	TC	MX		O		O	C/F	WP	H	0.13[a]
Brakel Olson	PP	E1		O		O	F	WP	H	0.31[a]
(1990)	PP	E1		O		O	F	WP	H	0.19[a]
Olson & Raffeld	TC	E2		TE	WR	CM	C	WP	H	1.12
(1987)	TC	E2		TE	WR	CM	C	WP	C	0.1
	TC	E2		TE	WR	CM	F	WP	H	0.02
	TC	E2		TE	WR	CM	F	WP	C	-0.9
Polio, Fleck, &	PP	E2	H	TE	WR	M	F	WP	GR	0.41[a]
Leder (1998)	TC	E2	H	TE	WR	M	F	WP	GR	-0.04[a]

Study	Design type	Language group	Level	Source of feedback	Mode of feedback	Type of feedback	Focus of feedback	Outcome type	Outcome focus	Effect size (or average effect size)
	PP	E2	H	TE	WR	M	F	WP	H	1.19
	PP	E2	H	TE	WR	M	F	WP	H	1.07
	PP	E2	H	TE	WR	M	F	WP	H	0.37
Sengupta (2000)	TC	E2		TE	OR	O		WP	H	0.68
	TC	E2		TE	OR	O		WP	H	0.93
Xiang (2004)	TC	E2		O	OR	CM	C/F	WP	H	0.7[a]
	TC	E2		O	OR	CM	C/F	WP	C	0.3[a]
	TC	E2		O	OR	CM	C/F	WP	GR	0.06[a]
Zimmerman &	TC	E1		TE	OR	CM	.	WP	H	0.53[a]
Kitsantas (2002)	TC	E1		TE	OR	CM	.	O		0.53[a]

Note. See Table 1 for a list of all variables, and a key to the variable codes.

[a] Averaged effect sizes.

Test of English as a Foreign Language
PO Box 6155
Princeton, NJ 08541-6155
USA

To obtain more information about TOEFL programs and services, use one of the following:

Phone: 1-877-863-3546
(US, US Territories*, and Canada)

1-609-771-7100
(all other locations)

E-mail: toefl@ets.org
Web site: www.ets.org/toefl

*America Samoa, Guam, Puerto Rico, and US Virgin Islands