Lang. Teach. (2010), **43:4**, 461–479 © Cambridge University Press 2010
doi:10.1017/S0261444810000200

Research Timeline

Research synthesis

John M. Norris and Lourdes Ortega University of Hawai'i at Mānoa, USA
jnorris@hawaii.edu, lortega@hawaii.edu

When in the late 1990s we set out to review the accumulated research on second language (L2) instructional effectiveness, it was with a sense – though no certainty – that a journey into the uncharted territory of systematic research synthesis would eventually prove worth the effort. In order to find a fitting methodology, and in the absence of any guidance within applied linguistics, we willingly delved into challenging techniques and debates from other disciplines. A few years later, we had to search long and hard to find enough synthesists working on language learning and teaching issues to warrant a book, but we did, and we saw that new effort as an important, if initial, step towards the adoption and adaptation of research synthesis, and particularly the subset of methods known as meta-analysis, to the problems of applied linguistics. Now, some ten years after our first publication on the topic, applications of meta-analysis have increased dramatically in the field, as this Timeline clearly shows. For some topics (e.g. interaction, corrective feedback), we have even seen several studies replicating and building on each other. We are pleased with these developments and welcome the efforts of the field to engage in rigorous and meaningful reviews of L2 research, with an eye towards cumulative explanation of key phenomena. At the same time, we also hope that researchers will avail themselves of the full potential of systematic research synthesis, which considerably transcends the narrower domain of meta-analysis and includes other quantitative as well as qualitative methods that more fully enable the task of synthesizing accumulated knowledge in the increasingly diverse research landscape of applied linguistics.

Clearly, systematic research synthesis is a recent introduction in applied linguistics, but it has a long and influential history in the social sciences. Accordingly, we have selected key publications and events that demarcate the landscape of synthesis not only in L2 research, but also in the emergence of core methodologies from psychology, education and health. For ease of viewing, we have shaded the non-L2 entries in the Timeline. Our selection of benchmark publications and events reflects six important themes, referenced in the last column of the Timeline as follows:

- **A** Empirical reviews subscribing to principles and techniques of synthesis, including meta-analyses (A.1), quantitative syntheses (A.2) and syntheses of qualitative research (A.3).
- **B** Reviews that draw on features of synthesis without adhering to the basic principles of the methodology.
- **C** Selected historical, methodological and conceptual discussions associated with the development of research synthesis.

https://doi.org/10.1017/S0261444810000200 Published online by Cambridge University Press

D Selected methodological guides for conducting research syntheses.
E Disciplinary activity associated with research synthesis (e.g. new editorial policies; creation of software, journals, associations or entities devoted to synthesis).
F Discussions of epistemological tensions and expansions in research synthesis, particularly in the service of synthesizing qualitative and mixed-methods research.

As with any other field, technical and epistemological developments continuously sharpen and renew the nature of what synthesists can do and the degree of reflexivity with which they do it. We hope applied linguists, particularly new generations, will cultivate a critical excitement towards synthesis and nurture a synthetic ethic for all of their research work. This Timeline chronicles the emergence of synthesis in applied linguistics as one approach to achieving this goal.

JOHN NORRIS is Associate Professor in the Department of Second Language Studies at the University of Hawai'i at Mānoa, USA, where he conducts research and teaches courses on second and foreign language assessment, evaluation, pedagogy and research methods. Recent books include *Toward useful program evaluation in college foreign language education* (co-edited with John McE. Davis, Castle Sinicrope and Yukiko Watanabe; National Foreign Language Resource Center, 2009), *Task-based language teaching: A reader* (co-editor with Kris Van den Branden and Martin Bygate; John Benjamins, 2009) and *Validity evaluation in language assessment* (Peter Lang, 2008).

LOURDES ORTEGA is Professor in the Department of Second Language Studies at the University of Hawai'i at Mānoa, USA, and Editor of *Language Learning*. Her research is in the areas of second language acquisition and instruction, L2 writing and research methods. Recent books are *The longitudinal study of advanced L2 capacities* (co-edited with Heidi Byrnes; Routledge, 2008) and *Understanding second language acquisition* (Hodder, 2009).

https://doi.org/10.1017/S0261444810000200 Published online by Cambridge University Press

https://doi.org/10.1017/S0261444810000200 Published online by Cambridge University Press

YEAR	REFERENCES	ANNOTATIONS	THEME
1904	Hunt, M. (1997). *How science takes stock: The story of meta-analysis*. New York: Russell Sage Foundation.	**Hunt** traces the pre-history of meta-analysis back to mathematician Karl Pearson, who in 1904 established the effectiveness of inoculation against typhoid fever by averaging correlations across several small-sample primary studies.	**C**
1968	*Language Teaching*, Cambridge University Press (ISSN: 0261-4448). URL: http://journals.cambridge.org/lta	***Language Teaching*** is the first journal in applied linguistics devoted to authoritative commissioned reviews and comprehensive bibliographies.	**E**
1976	Glass, G. V. (1976). Primary, secondary and meta-analysis of research. *Educational Researcher* 5.10, 3–8.	In his presidential address to the American Educational Research Association, **Glass** coined the term meta-analysis or 'analysis of analyses'. He contended that a tremendous increase in educational research demanded empirical integration of raw 'information' into orderly 'knowledge' and offered meta-analysis as the method to achieve just that.	**C** **A.1**
1977	Krashen, S. D. (1977). Some issues relating to the Monitor Model. In H. D. Brown, C. A. Yorio & R. H. Crymes (eds.), *On TESOL '77. Teaching and learning English as a second language: Trends in research and practice* Washington, DC: TESOL, 144–158.	**Krashen** employed several methods for aggregating and synthesizing studies of L2 morpheme accuracy orders. Based on rank-order correlations and implicational analyses, he concluded there was a shared 'natural order' in which these morphemes were mastered by L2 learners.	**B**
1977–1978	Smith, M. L. & G. V. Glass (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist* 32, 752–760. Schmidt, F. L. & J. E. Hunter (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology* 62, 529–540. Rosenthal, R. (1978). Combining results of independent studies. *Psychological Bulletin* 85, 185–193.	The specialized statistical literature on how to aggregate and combine primary data across studies quickly grew in the late 1970s. **Smith & Glass** was the first official meta-analysis. Independently, industrial-organizational psychologists **Schmidt** & **Hunter** and educational psychologist **Rosenthal** also contributed ground-breaking meta-analyses and methodological discussions.	**D** **A.1**
1978	Carver, R. P. (1978). The case against statistical significance testing. *Harvard Educational Review* 48, 378–399.	Criticisms against the abuse of statistical significance testing began appearing as early as the 1930s. This article by well-known educational psychologist **Carver** summarized the problems.	**C**

https://doi.org/10.1017/S0261444810000200 Published online by Cambridge University Press

YEAR	REFERENCES	ANNOTATIONS	THEME
1979	Rosenthal, R. (1979). The 'file drawer problem' and tolerance for null results. *Psychological Bulletin* 86, 638–641.	**Rosenthal** presented his well-known 'failsafe formula', which estimates the number of studies contributing zero effect sizes that would be needed to overturn a mean effect size found to be statistically significant.	**D**
1979	Krashen, S., M. H. Long & R. Scarcella (1979). Accounting for child–adult differences in second language rate and attainment. *TESOL Quarterly* 13, 573–582.	**Krashen et al.** reviewed the literature on the critical period hypothesis, examining findings across 18 studies by relative length of learning at the time of testing. They discovered that early starters outperformed late starters only for the five studies that compared participants after a minimum three to five years of learning. Their conclusion was that older is better initially, but that an early start affords an ultimate advantage if long-term outcomes are measured.	**B**
1980	Smith, M. L. (1980). Publication bias and meta-analysis. *Evaluation in Education* 4, 22–24.	Education scholar **Smith** was the first to use the term 'publication bias' in her paper, for the concern that studies with statistically significant findings have a higher probability of publication.	**D**
1980	*Annual Review of Applied Linguistics*, Cambridge University Press (ISSN: 0267-1905). URL: http://journals.cambridge.org/lta	***Annual Review of Applied Linguistics*** was founded by Robert Kaplan. It is devoted to commissioned state-of-the-art overviews and bibliographical updates.	**E**
1981	Glass, G. V., B. McGaw & M. L. Smith (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage.	**Glass et al.** was the first textbook explaining the methods of meta-analysis. The genre has since exploded, with numerous textbooks published since the mid-1980s.	**D**
1984	Light, R. & D. Pillemer (1984). *Summing up: The science of reviewing research*. Cambridge, MA: Harvard University Press.	**Light & Pillemer**'s classic text distinguishes itself for the authors' innovative thinking about synthesis (including meta-analysis and other synthetic approaches), even by today's standards, and for its readability. Highly cited, and highly recommended.	**D**

https://doi.org/10.1017/S0261444810000200 Published online by Cambridge University Press

1986	Slavin, R. E. (1986). Best evidence synthesis: An alternative to meta-analytic and traditional reviews. *Educational Researcher* 15.9, 5–11.	Educational psychologist **Slavin** proposed that only studies (published or unpublished) that meet the highest standards of methodological quality should be included in a meta-analysis.	**C** **D**
1987	Ragin, C. C. (1987). *The Comparative Method: Moving beyond qualitative and quantitative strategies*. Berkeley, CA: University of California Press.	Sociologist **Ragin** produced a set of techniques for the transformation of qualitative primary research into quantitative synthetic findings and relationships.	**F** **A.3**
1988	Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd edn.). Hillsdale, NJ: Lawrence Erlbaum.	A classic reading, key to understanding the relationship among magnitude, probability and power. **Cohen** also explains his rule-of-thumb benchmarks for interpreting *d*-values: small if $d = 0.2$ or just under a quarter of a standard deviation difference between conditions; medium if around $d = 0.5$ or half a standard deviation; and large if at or larger than $d = 0.8$ or over three-quarters of a standard deviation.	**D**
1988	Noblit, G. W. & R. D. Hare (1988). *Meta-ethnography: Synthesizing qualitative studies* Newbury Park, CA: Sage.	In this oft-cited book, **Noblit & Hare** synthesized three ethnographies of education and outlined an interpretive methodology known as 'meta-ethnography' or 'meta-synthesis'. The authors emphasized purposeful sampling, interpretation, subjectivity and induction as desirable standards for qualitative syntheses.	**F** **A.3**
1989	Auger, C. P. (1998). *Information sources in grey literature*. London: Bowker Saur.	**Auger** is a classic introductory text on the topic of grey literature – literature that is hard to retrieve because it is outside the commercial publisher circuits (also called 'gray' or 'fugitive' literature).	**D**
1991	Lazaraton, A. (1991). Power, effect size, and second language research. A researcher comments. *TESOL Quarterly* 25, 759–762.	**Lazaraton**'s brief commentary was probably the first direct discussion of effect sizes and the abuses of inferential statistics published in an applied linguistics journal.	**C**

https://doi.org/10.1017/S0261444810000200 Published online by Cambridge University Press

YEAR	REFERENCES	ANNOTATIONS	THEME
1993	Cochrane Collaboration: www.cochrane.org Centre for Reviews and Dissemination (CRD): www.york.ac. uk/inst/crd/index.htm Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre): eppi.ioe.ac.uk/cms	Focusing on healthcare issues, the **Cochrane Collaboration** and **CRD** are two not-for-profit organizations created the same year, devoted to producing, disseminating and policing systematic reviews. Similarly, the **EPPI-Centre** focuses on syntheses in social science and public policy.	**E**
1993	Lipsey, M. W. & D. B. Wilson (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist* 48, 1181–1209.	In this oft-cited meta-analysis of meta-analyses, public policy researchers **Lipsey & Wilson** examined, among other things, the extent to which methodological quality impacted results. Based on 302 meta-analyses, they demonstrated that some, but not most, sources of differing research quality made a systematic difference on the results reported.	**C** **A.1**
1994	Cooper, H. & L. V. Hedges (eds.) (1994). *The handbook of research synthesis*. New York: Russell Sage Foundation.	Educational statisticians **Cooper** & **Hedges** were funded by the Russell Sage Foundation to produce this seminal effort, in which experts covered main topics associated with the methodology. Despite the generic title, it focuses primarily on meta-analysis. Advanced, technical reading.	**C** **D**
1994	Editor's statement. *Language Learning* 44.1, p. 3	**Cumming**, then editor of *Language Learning*, published an editorial in which he announced the journal's creation of a special Review Article section. Increasingly, reviews published in *Language Learning* have been meta-analyses.	**E**
1994	Thomas, M. (1994). Assessment of L2 proficiency in second language acquisition research. *Language Learning* 44, 307–336.	In this early synthesis, **Thomas** examined the measurement and reporting of proficiency across 157 empirical L2 studies published in four journals between 1988 and 1992. Four categories of measures emerged: institutional status (40%), standardized tests (22%), impressionist judgments (21%) and in-house tests (14%). The findings demonstrate the power of synthesis for generating insights that can challenge methodological practices and theoretical assumptions.	**A.2**

https://doi.org/10.1017/S0261444810000200 Published online by Cambridge University Press

1996	Thompson, B. (1996). AERA editorial policies regarding statistical significance testing: Three suggested reforms. *Educational Researcher* 25.2, 26–30.	Educational psychologist **Thompson** advocates effect sizes, proposing that authors: (a) use more precise language when talking about statistical significance, (b) report and capitalize on effect sizes for all interpretations and (c) assess the replicability of results.	**C**
1997	Harlow, L., S. Mulaik & J. Steiger (eds.) (1997). *What if there were no significance tests?* Mahwah, NJ: Lawrence Erlbaum.	**Harlow et al.**'s edited collection quickly became a classic in the movement to reform statistical inference and reporting in the social sciences.	**C**
1997	Sahari, M. (1997). Elaboration as a text-processing strategy: A meta-analytic review. *RELC Journal* 28, 15–27.	**Sahari** is the first meta-analysis proper published in an applied linguistics journal, although it investigated the topic based on L1 studies.	**A.1**
1998	Cochrane Qualitative Research Methods Group: www.joannabriggs.edu.au/cqrmg	In the late 1990s, researchers mobilized efforts to advance methods for the systematic synthesis of qualitative research. The **CQRM group** was the first to be established; in 2001 the Campbell Collaboration also saw the creation of a similar group, the Implementation Process Methods Group.	**E** **F**
1998–1999	*Review of Educational Research* 68.4 and 69.1.	***RER***, the flagship journal of the American Educational Research Association, featured a series of special papers that discussed qualitative reviewing by distinguished contributors such as Eisenhart, Schwandt and Lather.	**F**
1998	Ross, S. (1998). Self-assessment in second language testing: A meta-analysis and analysis of experiential factors. *Language Testing* 15, 1–20.	**Ross** reviewed 10 studies and aggregated 60 correlations which yielded an average $r = 0.63$, indicating that the mean shared variance between a self-assessment and some direct measure of proficiency in these studies was 40%.	**A.1**
1998	Wolfe-Quintero, K., S. Inagaki & H.-Y. Kim (1998). *Second language development in writing: Measures of fluency, accuracy, and complexity*. Honolulu, HI: University of Hawai'i, Second Language Teaching and Curriculum Center.	**Wolfe-Quintero et al.** sought to identify the best indices for measuring L2 complexity, accuracy and fluency, based on 39 studies of essay writing by high school and college learners. The main method employed was vote-counting, deployed in tabular	**B**

https://doi.org/10.1017/S0261444810000200 Published online by Cambridge University Press

YEAR	REFERENCES	ANNOTATIONS	THEME
		summaries of varying degrees of significance yielded by comparisons of measures and proficiencies across studies. The vote-counting approach is generally not accepted as a method for systematic synthesis, due to dependence on sample sizes, significance tests and likely publication biases.	
1999	Wilkinson, L. & Task Force on Statistical Inference (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist* 54, 594–604.	A committee of the Board of Scientific Affairs of the American Psychological Association produced this policy statement in which they strongly endorsed statistical reforms, including the reporting of effect sizes and confidence internals for all findings (regardless of statistical significance).	**E**
1999	Krashen, S. (1999). Seeking a role for grammar: A review of some recent studies. *Foreign Language Annals* 32, 245–257.	Although the study did not adhere to principles of systematic synthesis (e.g. exhaustive sampling or study eligibility criteria), **Krashen** was the first to employ effect sizes as the primary means to ascertain whether L2 instruction is effective.	**B**
2000	Campbell Collaboration: www.campbellcollaboration.org	The **Campbell Collaboration** focuses on reviews of social interventions in education, criminology and social welfare.	**E**
2000	Comprehensive Meta-Analysis: www.meta-analysis.com	This **meta-analysis software** was funded by the US National Institutes of Health and developed by US and UK experts, many of them authors of the references cited in this Timeline. Although not the first or only software for meta-analysis, it is the most sophisticated and complete.	**D** **E**
2000	Editor's statement. *Language Learning* 50.3, xi–xiii.	**N. Ellis**, then editor of *Language Learning*, instituted a new policy that required authors to report effect sizes, 'at least for the major statistical contrasts they report' (p. xii). Robert DeKeyser reiterated the policy in 2007. No other applied linguistics journal requires the reporting of effect sizes, although *TESOL Quarterly* mentions it as desirable in C. A. Chapelle & P. A. Duff (eds.) (2003), Some guidelines for conducting quantitative and qualitative research in TESOL. *TESOL Quarterly* 37, 157–178.	**E**

https://doi.org/10.1017/S0261444810000200 Published online by Cambridge University Press

2000	Norris, J. M. & L. Ortega (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning* 50, 417–528.	**Norris & Ortega** conducted the first comprehensive exemplar of meta-analysis in applied linguistics. They synthesized research practices (based on 77 reports) and meta-analyzed the effectiveness of L2 instructional treatments (based on 49 unique sample studies). The mean effect size for instruction was $d = 0.96$, a clear advantage over no instruction or sheer exposure. Reliable advantages in magnitude were found for explicit over implicit techniques. Nine recommendations were offered for the improvement of methodological quality. The 112-page article provides a detailed guide to the application of synthesis and meta-analysis.	**A.1** **D**
2001	Lipsey, M. W. & D. B. Wilson (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.	**Lipsey & Wilson**'s textbook distinguishes itself in its detailed but accessible approach. With many useful appendices and clear discussions of technical issues that arise over the entire process of meta-analysis, it is perhaps the best introduction to conducting meta-analysis.	**D**
2001	Goldschneider, J. & R. M. DeKeyser (2001). Explaining the 'natural order of L2 morpheme acquisition' in English: A meta-analysis of multiple determinants. *Language Learning* 51, 1–50.	Inspecting the findings across 12 studies dated 1973–1996, **Goldschneider & DeKeyser** sought to explain the morpheme accuracy order (KRASHEN 1977). Extracted from each study were accuracy scores, expressed as group mean performance percentages weighted by sample size, each associated with 6 particular morphemes which, in turn, were rated by the meta-analysts for six predictive factors (e.g. perceptual salience, frequency, etc.). A multiple regression analysis revealed that each predictive variable explained 16–36% of the shared variance on its own, and that all combined predicted 71% of the shared variance with accuracy scores. The study is unique in using meta-analysis not for the sake of summarizing previous findings, but with the purpose of gleaning new theoretical knowledge.	**A.1**

https://doi.org/10.1017/S0261444810000200 Published online by Cambridge University Press

YEAR	REFERENCES	ANNOTATIONS	THEME
2002	What Works Clearinghouse: http://ies.ed.gov/ncee/wwc	The US Department of Education's Institute of Education Sciences established the **WWC** to police and promote the quality of syntheses and reviews about education. The WWC initiative has been found successful and useful by many users. It has also attracted heavy criticism from qualitative researchers in the educational research community for its promotion of one-sided forms of knowledge about education and its emphasis on a narrow understanding of 'scientifically based research'.	**E**
2003	Finfgeld, D. L. (2003). Metasynthesis: The state of the art – so far. *Qualitative Health Research* 13, 893–904.	NOBLIT & HARE'S (1988) meta-synthesis method, even though it originated in education, has proven very useful among health synthesists. **Finfgeld** offers a review of its diverse applications in this field for the period of 1994–2002.	**F**
2003	Masgoret, A.-M. & R. C. Gardner (2003). Attitudes, motivation, and second language learning: A meta-analysis of studies conducted by Gardner and associates. *Language Learning* 53, 123–163.	**Masgoret & Gardner** synthesized the relationship of L2 achievement with the three constructs of motivation, attitude and orientation, as measured by the widely used Attitude/Motivation Test Battery (developed by Gardner). The authors concluded that the relationship of L2 achievement to motivation was 'more than medium'. This meta-analysis presents high internal validity (pooled observations for more than 10,000 participants) and low external validity (all studies included were carried out by Gardner and associates).	**A.1**
2003	Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics* 24, 492–518.	**Ortega** demonstrated the potential of syntheses to generate benchmarks for the interpretation of future research findings. Only two measures of complexity turned out to be consistently employed across 21 cross-sectional and 6 longitudinal studies: the mean length of T-unit (MLTU) and the mean number of clauses per T-unit (C/TU). Differences of 2 words or higher in MLTU or 0.20 or higher in C/TU were proposed as critical effect sizes, i.e. in relation to meaningful differences. The study made innovative use of stem-and-leaf plots.	**A.2**

https://doi.org/10.1017/S0261444810000200 Published online by Cambridge University Press

2003	Zhao, Y. (2003). Recent developments in technology and language learning: A literature review and meta-analysis. *CALICO Journal* 21, 7–27.	**Zhao** synthesized main themes investigated in 156 studies published across 22 refereed journals from 1997 to 2001 and meta-analyzed the relative effectiveness of technology-supported L2 interventions in 9 of them.	**A.1**
2005	Rothstein, H. R., A. J. Sutton & M. Borenstein (eds.) (2005). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. Chichester: Wiley.	The unique collection edited by **Rothstein et al.** reviews technical and sociological aspects of publication bias. Useful chapter preview/summaries and bibliographical appendix.	**C** **D**
2006	Dixon-Woods, M., S. Bonas, A. Booth, D. R. Jones, T. Miller, A. J. Sutton, R. Shaw, J. A. Smith & B. Young (2006). How can systematic reviews incorporate qualitative research? A critical perspective. *Qualitative Research* 6, 27–44.	The UK-based team led by health scholar **Dixon-Woods et al.** discusses Critical Interpretive Synthesis as an epistemologically congruent approach to the synthesis of qualitative primary research.	**F**
2006	Petticrew, M. & H. Roberts (2006). *Systematic reviews in the social sciences: A practical guide*. Oxford: Blackwell.	**Petticrew & Roberts**' textbook is refreshing in that it introduces readers to methods for synthesizing both quantitative and qualitative studies. Oriented towards health research, but not as exclusively as several other existing handbooks and textbooks on qualitative and quantitative synthesis methods from the health disciplines.	**D** **F**
2006	Society for Research Synthesis Methodology: www.srsm.org	The **SRSM** was founded and convened its first annual meeting this year in Cambridge, UK, holding annual conferences since.	**E**
2006	Dinsmore, T. H. (2006). Principles, parameters, and SLA: A retrospective meta-analytic investigation into adult L2 learners' access to Universal Grammar. In Norris & Ortega (eds.), 53–90.	Addressing the broad question of whether L2 learners perform similarly to or differently from native speakers, **Dinsmore** meta-analyzed 16 studies dated 1981–1999. On average, the L2 learners were different from their native speaking comparisons by well over a standard deviation unit, $d = 1.25$. Interesting features were the detailing of reasons for excluding studies, and the use of homogeneity tests, which indicated that some moderator variables were probably causing some of the variance.	**A.1**

https://doi.org/10.1017/S0261444810000200 Published online by Cambridge University Press

YEAR	REFERENCES	ANNOTATIONS	THEME
2006	Indefrey, P. (2006). A meta-analysis of hemodynamic studies on first and second language processing: Which suggested differences can we trust and what do they mean? *Language Learning* 56 Supplement 1, 279–304.	Despite the word 'meta-analysis' in the title, **Indefrey** reported on a narrative synthesis of twenty-four studies of blood flow and brain activation during bilingual language processing, to discern possible differences in the areas that are activated during L1 vs. L2 processing among bilinguals. The author concluded that the areas of activation are the same for L2 and L1, but there is a tendency towards stronger activation in the L2 when engaged in some but not all tasks, and depending on onset of learning, relative proficiency and exposure.	**B**
2006	Jeon, E. H. & T. Kaya (2006). Effects of L2 instruction on interlanguage pragmatic development: A meta-analysis. In Norris & Ortega (eds.), 165–211.	Interested in assessing the effectiveness of L2 instruction on pragmatics, **Jeon & Kaya** meta-analyzed 13 studies dated from 1984 to 2003. They uncovered a mean effect size of $d = 0.59$ for treatment/control contrasts, accompanied by very large variability across studies. Interesting are the analysis of pragmatic instructional features and the inclusion of 'grey literature'. The synthesis also found that many pragmatic treatments are rather explicit and last more than five hours, in sharp contrast with instruction of grammar, which tends to include at least some implicit treatment and is typically brief, less than one hour (see NORRIS & ORTEGA 2000).	**A.1**
2006	Keck, C. M., G. Iberri-Shea, N. Tracy-Ventura & S. Wa-Mbaleka (2006). Investigating the empirical link between task-based interaction and acquisition: A meta-analysis. In Norris & Ortega (eds.), 91–131.	Responding to repeated doubts that interaction has an effect on acquisition, and not just performance, **Keck et al.** meta-analyzed 14 studies that featured pre-post tests and control group designs, dated 1994–2003. They found a reliable average effect of $d = 0.92$ for the groups that had engaged in task-based interaction focusing on grammar or lexis. A novel theoretical contribution was the inclusion of task essentialness as a moderator variable not found in any primary study but superimposed by the meta-analysts. As it turned out, task-essentialness led to noticeable learning increments on delayed post-tests.	**A.1**

https://doi.org/10.1017/S0261444810000200 Published online by Cambridge University Press

2006	Norris, J. M. & L. Ortega (eds.) (2006). *Synthesizing research on language learning and teaching*. Amsterdam: John Benjamins.	**Norris & Ortega** offered the first collection devoted to synthesis and meta-analysis in applied linguistics. It included five meta-analyses and two qualitative syntheses, each on a different substantive issue. An opening chapter by the co-editors laid out a space in the field for a synthetic research ethic, cautioning against an excessively technicist, cookie-cutter approach to meta-analysis, and inviting researchers to engage in diverse forms of synthesis, including qualitative synthesis.	**D F**
2006	Russell, J. & N. Spada (2006). The effectiveness of corrective feedback for the acquisition of L2 grammar: A meta-analysis of the research. In Norris & Ortega (eds.), 133–164.	**Russell & Spada** examined the effects of L2 error correction across 15 studies published between 1988 and 2003. The mean effect size was $d = 1.16$. Interesting was the weighting of effect sizes by sample size, the effort to inspect the impact of methodological quality via sensitivity analyses, and the full reporting of results for study coding.	**A.1**
2006	Taylor, A. M. (2006). The effects of CALL versus traditional L1 glosses on L2 reading comprehension. *CALICO Journal* 23, 309–318.	**Taylor** meta-analyzed 12 studies published between 1989 and 1999. While the mean effect of using L1 glosses during computer-mediated L2 reading was large ($g = 1.09$), the effect of L1 glosses made available during traditional (paper) L2 reading was small ($g = 0.39$). Although the difference was reliable, the small sample size calls for caution. Noteworthy is the use of a forest plot, which helps assess graphically the variability across individual study results.	**A.1**
2006	Taylor, A., J. R. Stevens & J. W. Asher (2006). The effects of explicit reading strategy training on L2 reading comprehension: A meta-analysis. In Norris & Ortega (eds.), 213–244.	**Taylor et al.** investigated how effective it is for L2 learners to be taught explicitly a range of strategies that are known to support reading comprehension, as measured by recall protocols in most of these studies. The mean effect size of $g = 0.54$ was considered medium but important, in that 68% of average L2 readers will comprehend better with the strategy training than average readers without it. Of particular interest was the inclusion of 'grey literature' and the assessment of publication bias via a sensitivity analysis.	**A.1**

https://doi.org/10.1017/S0261444810000200 Published online by Cambridge University Press

YEAR	REFERENCES	ANNOTATIONS	THEME
2006	Téllez, K. & H. C. Waxman (2006). A meta-synthesis of qualitative research on effective teaching practices for English Language Learners. In Norris & Ortega (eds.), 245–277.	**Téllez & Waxman** offer the only meta-synthesis of qualitative studies to date in applied linguistics. The authors synthesized 25 studies of high-quality instructional practices for English Language Learners in K-12 schools, using NOBLIT & HARE's (1988) methodology. In the interpretive, rather than aggregative, spirit of meta-synthesis, five themes were distilled to account for best practices, including communitarian teaching and dialogic teacher–student interactions.	**A.3** **F**
2006	Thomas, M. (2006). Research synthesis and historiography: The case of assessment of second language proficiency. In Norris & Ortega (eds.), 279–298.	**Thomas** theorized synthesis and compared it to historiography by presenting the results of a study of changes in the measurement of proficiency in L2 research, twelve years after THOMAS (1994). She reports subtle changes, including a concurrent trend to downplay proficiency as a variable (among L2 researchers from formal linguistic approaches) and to employ multiple and increasingly sophisticated techniques to determine participants' proficiency (among other L2 researchers).	**A.2** **F**
2007	Krashen, S. (2007). Extensive reading in English as foreign language by adolescents and young adults: A meta-analysis. *The International Journal of Foreign Language Teaching* 3.2, 23–29.	In this review, as in KRASHEN 1999, **Krashen** did not adhere to certain principles of synthesis, particularly for the selection of studies. However, he employed effect sizes and other quantitative techniques in order to glean evidence across 13 studies about the effectiveness of extensive reading for English as a foreign language learners. Krashen concluded that 'effect sizes were uniformly positive and typically quite impressive' (p. 27).	**B**
2007	Mackey, A. & J. M. Goo (2007). Interaction research in SLA: A meta-analysis and research synthesis. In A. Mackey (ed.), *Conversational interaction in second language acquisition: A collection of empirical studies*. New York: Oxford University Press, 407–452.	**Mackey & Goo** replicated and extended the meta-analysis by KECK ET AL. (2006), including 12 new studies. With the increase in studies, the mean effect ($d = 0.75$) was more trustworthy, although a concern regarding external validity is that half ($n = 6$) of the studies since 2003 were published in the same book where the meta-analysis appeared. The study provides a good example	**A.1**

https://doi.org/10.1017/S0261444810000200 Published online by Cambridge University Press

		of synthetic findings making important theoretical advances, in that a dissymmetry was uncovered between the magnitude of lexical learning, which was larger than for grammar but diminished slightly on delayed post-tests, and the effect for grammatical learning, which was initially of medium size but became large after a time lag.	
2007	Norris, J. M. & L. Ortega (2007). The future of research synthesis in applied linguistics: Beyond art or science. *TESOL Quarterly* 41, 805–815.	**Norris & Ortega** discussed the benefits and challenges of the methodology as it has begun to be practiced in applied linguistics. They called for attention to epistemologically diverse methods (e.g. synthesizing qualitative research) so as to be responsive to intellectual developments in the field.	**C F**
2007	Truscott, J. (2007). The effect of error correction on learners' ability to write accurately. *Journal of Second Language Writing* 16, 255–272.	**Truscott** contributed this synthesis to the debate of whether correcting grammar errors in L2 writing is beneficial. The author based his conclusions on two sets of effect sizes drawn from 6 'controlled experiments' and from 6 'absolute gains' studies. On the one hand, selecting only 12 studies out of a much larger domain could be conceived of as best evidence synthesis. On the other hand, readers may question the reasons for excluding other studies. For highly controversial topics, such as error correction or bilingual education, no synthetic or meta-analytic study is likely to solve these dilemmas, since what counts as wise selectivity versus bias is in the eye of the beholder (or in the vested interests of the experts).	**B**
2008	**APA Publications and Communications Board Working Group on Journal Article Reporting Standards** (2008). Reporting standards for research in psychology: Why do we need them? What might they be? *American Psychologist* 63.9, 839–851.	This article delineated optimal requirements for research reporting in psychology and informed the 6th edition of the *APA Manual*. Effect sizes, confidence intervals and power are noted as required reporting elements in primary quantitative studies. Detailed reporting standards for meta-analysis are also offered. The JARS Group was chaired by meta-analyst Harris Cooper.	**E**

https://doi.org/10.1017/S0261444810000200 Published online by Cambridge University Press

YEAR	REFERENCES	ANNOTATIONS	THEME
2008	Abraham, L. B. (2008). Computer-mediated glosses in second language reading comprehension and vocabulary learning: A meta-analysis. *Computer Assisted Language Learning* 21, 199–226.	**Abraham** meta-analyzed 11 studies published between 1994 and 2007, out of an initial pool of 82. Computer glosses during L2 online reading led to a medium and imprecise advantage for reading comprehension ($d = 0.73$, with very wide confidence intervals) but to a large and robust advantage for vocabulary learning ($d = 1.40$) that was sustained over delayed post-tests. Interesting was the reported use of the software COMPREHENSIVE META-ANALYSIS.	**A.1**
2008	Lee, S.-K. & H. T. Huang (2008). Visual input enhancement and grammar learning: A meta-analytic review. *Studies in Second Language Acquisition* 30, 307–331.	Drawing on 16 studies of input enhancement, **Lee & Huang** synthesized research practices and meta-analyzed quantitative results. The main findings were a large variability across studies and overall small effects, and the use of comparison groups that experience input flood, rather than true controls with no exposure. Interesting was the use of sensitivity analysis that yielded tentative evidence of publication bias in this domain.	**A.1**
2008	Wa-Mbaleka, S. (2008). *A meta-analysis investigating the effects of reading on second language vocabulary learning*. Saarbrücken, Germany: VDM.	**Wa-Mbaleka** synthesized and meta-analyzed 44 studies of vocabulary learning during L2 reading published between 1981 and 2005. The overall effectiveness of reading was $d = 1.43$, with large between-study variability and also a wide confidence interval yet a lower boundary of $d = 0.75$, suggesting a robust effect despite impreciseness due to measurement error. Step-by-step explanations of the methodology are given.	**A.1**
2009	Cooper, H. M., L. V. Hedges & J. Valentine (eds.), *The handbook of research synthesis and meta-analysis* (2nd edn.). New York: The Russell Sage Foundation.	This second edition of the seminal 1994 publication by COOPER & HEDGES is an updated and strengthened version. The new title makes explicit the strong focus on quantitative meta-analysis and its statistical techniques.	**C** **D**
2009	Suri, H. & D. Clarke (2009). Advancements in research synthesis methods: From a methodologically inclusive perspective. *Review of Educational Research* 79, 395–430.	Australian educational researchers **Suri & Clarke** surveyed in depth the history of qualitative methods for synthesizing qualitative research, countering the predominant acceptance of a quantitative, neo-positivist perspective.	**A.3** **F**

https://doi.org/10.1017/S0261444810000200 Published online by Cambridge University Press

2009	In'nami, Y. & R. Koizumi (2009). A meta-analysis of multitrait–multimethod studies in language testing research: Focus on language ability and Chapelle's (1998) construct definition and interpretation. *Language Testing* 26, 219–244.	**In'nami & Koizumi** employed meta-analysis to investigate whether multiple-choice versus open-ended response formats lead to measurement differences across 37 studies, focusing on tests of L1 reading, L2 reading, or L2 listening. Fifteen moderator variables were also examined but the number of accumulated observations turned out to be insufficient. The authors concluded that clear effects were at work for multiple-choice testing in L1 reading and L2 listening tests, leading to higher scores than open-ended formats, most noticeably for L2 listening.	**A.1**
2009	Taylor, A. M. (2009). CALL-based versus paper-based glosses: Is there a difference in reading comprehension? *CALICO Journal* 27, 147–160.	**Taylor** meta-analyzed 32 studies that featured the use of L1 or L2 glosses during computer-mediated L2 reading (average $g = 0.92$) versus traditional (paper) L2 reading (average $g = 0.43$). The difference in mean benefits for L2 reading comprehension was statistically significant, as shown in a homogeneity test.	**A.1**
2010	*Research Synthesis Methods* (doi10.1002/jrsm.0), published by Wiley-Blackwell.	***RSM*** is the new official journal of the Society for Research Synthesis Methods. Mark Lipsey and Christopher Schmid are the editors. It will publish research articles, as well as tutorials, reviews and bibliographies. Qualitative synthesis is among the topics listed.	**E**
2010	In'nami, Y. & R. Koizumi (2010). Database selection guidelines for meta-analysis in applied linguistics. *TESOL Quarterly* 44, 169–184.	**In'nami & Koizumi** discuss for the first time in applied linguistics issues surrounding the literature identification and retrieval stage of systematic synthesis.	**C**
2010	Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning* 60, 309–365.	Meta-analyzing 33 studies of corrective feedback (in laboratory, classroom and small group settings) published between 1988 and 2007, **Li** included several methodological innovations, from the examination of outliers to the comparison of results from fixed versus random effects models. Particularly valuable was the effort to address publication bias via the inclusion of 11 unpublished dissertations. The effectiveness of oral corrective feedback was $d = 0.61$ with a narrow confidence interval of 0.51 to 0.71.	**A.1**

https://doi.org/10.1017/S0261444810000200 Published online by Cambridge University Press

YEAR	REFERENCES	ANNOTATIONS	THEME
2010	Lyster, R. & K. Saito (2010). Oral feedback in classroom SLA: A meta-analysis. *Studies in Second Language Acquisition* 32, 265–302.	Also about oral corrective feedback, **Lyster & Saito**'s meta-analysis focused on classroom studies only (n = 15). The mean effect size for classroom corrective feedback was $d = 0.74$, with the confidence interval from 0.58 to 0.90. The authors usefully discuss emergent patterns from this and previous syntheses, such as lagged learning effects, wider use of free response formats than in other areas of L2 instructional research and concomitant effects of outcome measures.	**A.1**
2010	Ortega, L. (2010). Research synthesis. In B. Paltridge & A. Phakiti (eds.), *Companion to research methods in applied linguistics*. London: Continuum, 111–126.	**Ortega** introduces readers to the main features of synthesis. The discussion includes phases for conducting a synthesis as well as issues of validity, trustworthiness and ethics.	**D** **F**
2010	Oswald, F. L. & L. Plonsky (2010). Meta-analysis in second language research: Choices and challenges. *Annual Review of Applied Linguistics* 30, 85–110.	**Oswald & Plonsky** chart the territory of meta-analysis in applied linguistics, connecting it in useful ways with the wider and more technical meta-analysis literature in the social sciences (note that the first author was a member of the team that developed the software COMPREHENSIVE META-ANALYSIS, 2000). Guidance is provided on how and when to use (or not) certain features, such as homogeneity tests, the failsafe formula or fixed and random models. The authors also offer a useful methodological commentary on 27 meta-analyses in the field and advocate the development of discipline-specific benchmarks for interpreting *d*-values in L2 research.	**C** **D**
2010	Spada, N. & Y. Tomita (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning* 60, 263–308.	Based on effect sizes from 41 unique study samples published between 1990 and 2006, **Spada & Tomita** examined whether explicit or implicit instruction leads to varying gains depending on whether they target 'simple' or 'complex' structures. Overall, explicit treatments were more effective for either type of structure ($d = 0.73$ for simple and $d = 0.88$ for complex vs. 0.33 and 0.39 under implicit conditions) and sustained over time. The authors	**A.1**

https://doi.org/10.1017/S0261444810000200 Published online by Cambridge University Press

	note positive improvements in the domain, with more studies employing outcome measures that tap implicit L2 knowledge via free response formats (e.g. essays, oral narratives). As in other meta-analyses (e.g. MACKEY & GOO 2006; LI 2010; LYSTER & SAITO 2010), free response measures exhibited larger effects, a pattern contrary to that observed by NORRIS & ORTEGA (2000) in the research representative of the 1990s.	
Plonsky, L. (in press). The effectiveness of second language strategy instruction: A meta-analysis. *Language Learning*.	**Plonsky** meta-analyzed 59 studies investigating L2 strategy training. The main finding was a medium but reliable effect between $d = 0.44$ and $d = 0.53$ (upper and lower boundary of confidence internals). Noteworthy are the detailed methodological discussions, the use of the funnel plot to estimate publication bias, and the assessment of the impact of methodological quality on the reported meta-analytical results.	**A.1**