PLenary Speech

Coming of age: the past, present, and future of quantitative SLA research

Susan Gass1,2*, Shawn Loewen2 and Luke Plonsky3

1Southeast University, Nanjing, Jiangsu Province, China, 2Michigan State University, East Lansing, United States and 3Northern Arizona University, Flagstaff, United States
*Corresponding author. E-mail: susangass@gmail.com

Abstract
First, we trace the history of second language acquisition (SLA) from early stages in the mid-twentieth century to today. We next consider the status of the field in today’s research world with a particular focus on all aspects of methodology and, finally, we take a look at the future and discuss issues related to scientific rigor in light of Open Science.

Introduction
The field of second language (L2) research has undergone many changes in orientation as well as in research design and elicitation techniques. In this paper we trace some of the history of SLA research. We do this not from any particular research tradition; rather, we consider the changing criteria for research in the field and look at those changing criteria as they parallel the establishment of the field as a discipline in its own right. More specifically, the focus in the paper is on the development of rigor in the field. In meeting this objective, we limit ourselves to quantitative research given our own research experiences. We first consider how the development and establishment of the field parallels an increased focus on research methods and analysis. This will be followed by a discussion of indicators of current emphasis on rigor. We will close with a look at the future and how the field has positioned itself to move forward to ensure that our research results are ‘robust and unbiased,’ thereby ensuring the stability and ultimate value of the field. In sum, we highlight where we have come from, where we are now, and what the future directions in the field are with regard to increased scientific responsibility. We recognize that new ideas require new tools for elicitation and for analysis. Regardless of what those new tools are or will be in the future, and regardless of theoretical orientation, the field can only thrive and develop if our research demonstrates rigor. And, it is our strong belief that rigor in research crosses theoretical boundaries.

Rigor in scientific disciplines
Scientific rigor is the sine qua non in all areas of scientific inquiry. In a discussion of grant applications, the National Institutes of Health1 make the following comment about rigor:

the strict application of the scientific method to ensure robust and unbiased experimental design, methodology, analysis, interpretation, and reporting of results. (https://nexus.od.nih.gov/all/2016/01/28/scientific-rigor-in-nih-grant-applications/)

1Revised version of a plenary presented at the Montpelier conference on Research Methodology in the Field of Second Language Acquisition and Learning (30 May 2018).

© Cambridge University Press 2020
They go on to acknowledge that different disciplines will have their ‘own set of best practices or standards to achieve rigor.’ Nonetheless, whatever particular approach a field assumes as best practice, the ultimate goal of research in any field is to understand how a particular research design and method ‘will achieve robust and unbiased results.’ Among the basic criteria are the use of appropriate statistics, estimates of needed sample size, and designs that yield unambiguous results and that minimize potential threats to validity. In particular, ‘[r]obust and credible results are those obtained with methods specifically designed to avoid bias, such as blinding, randomization, and prospectively defined exclusion/inclusion criteria, to name a few.’ These issues are being debated and have taken center stage in all areas of scientific inquiry, the field of SLA being no exception.

The history of the field is such that the recognition of SLA as something separate from other disciplines and, in particular, separate from language teaching has gone hand-in-hand with developing rigor as well as with a focus on ensuring that we are skeptically and scientifically investigating constructs of interest. To understand our argument, it is first important to establish SLA as a discipline that covers a body of knowledge recognizing and adhering to principles of scientific rigor. As noted by Gass, Fleck, Leder, and Svetics (1998) ‘respect for the field…can come only through sound scientific progress’ (p. 407).

In an important editorial statement from Byrnes (2013), the then editor of Modern Language Journal raised the issue of the ‘methodological turn.’ In her words:

Methodological issues inherently merit a certain level of attention inasmuch as they assure the quality of our work. But it appears that at this point in the development of applied linguistics, they demand a kind of professional scrutiny that goes directly to the core of what we do and what we know and what we can tell our publics that we know—and not only how we do it. (p. 825)

Along similar lines, Plonsky (2011) begins his dissertation with two assumptions: ‘Assumption 1: Study quality matters. If the means by which researchers design, carry out, and report on their studies lack in rigor or transparency, theory and practice are likely to be misguided or at least decelerated. Assumption 2 is an implication of Assumption 1: Quality should be measured rather than assumed’ (Abstract). We return to current practices and issues related to study quality later in this paper.

The development of the discipline: a brief synopsis

As an initial step, it is important to understand the development of the field and the establishment of SLA as an independent discipline. We will not go into detail, nor do we consider relevant to our argument what the beginning of the field is (see Gass et al., 1998 and Thomas, 1998 for discussions on this issue). For purposes of this discussion, we take a somewhat arbitrary starting point of the 1960s or 1970s. In Selinker (1972) we saw some of the early arguments regarding appropriate data on which to base conclusions about L2 learning (see Gass & Polio, 2014 for further elaboration). As further evidence of the field’s founding, we might also point to events such as the emergence of doctoral programs and dedicated journals (e.g., Studies in Second Language Acquisition) and conferences such as the Second Language Research Forum (SLRF), which came into being at about this time, as we discuss below.

Whatever we take as the origins of the field, SLA grew out of concerns related to language teaching, a discipline with an unquestionably long history. To wit, the title of Kelly’s (1969) book is: 25 centuries of language teaching: 500 BC[E]–1969. His book treats topics of pedagogical interest, but also pays some attention to language learning when he talks about what one needs to pay attention to when designing a language course—the nature of language and what constitutes language learning.

Even prior to our arbitrary starting point of the 1960s/1970s, Lado (1957) and Fries in the foreword to Lado’s book recognized the importance of understanding principles of language learning on which language teaching materials should be based: ‘[t]he struggle to apply to the problems of foreign language learning the new views of language arising out of “structural” analyses has served to shift the focus of attention from methods and techniques of teaching to the basis upon which to build these materials’ (Fries, 1957, no page indicated).
What we see is that the ultimate purpose of research into language learning in these earlier times was to improve pedagogical approaches, tools, and methodologies. SLA was therefore subservient to the discipline of language teaching in the sense that its existence depended on its value to another field.

Even when the overt emphasis was on language learning, there was an underlying emphasis on language teaching. In the late 1960s, the oldest journal dedicated to language learning (Language Learning) published articles that were primarily pedagogical in orientation. Based on data from Gass and Polio (2014), most papers in the five-year period from 1967 to 1972 dealt with language pedagogy (40%) and only a small percentage (approximately 10%) was based on experimental data analysis of language learning. In the six years following, there is a noticeable shift in orientation with data analysis papers (approximately 46%) far exceeding pedagogically oriented papers (15%). As reported in Gass and Polio (2014), this shift is likely due to the increased focus on learner data rather than on teaching. They further argued that the shift in focus was likely due to the publication of Selinker’s well-known 1972 Interlanguage paper. As was also reported in Gass and Polio (p. 152), at that time, there was considerable discussion leading up to Selinker’s publication (particularly in Edinburgh with S. Pit Corder, Larry Selinker, Elaine Tarone, and others) that stimulated a serious look at learner data. In addition, perhaps a turning point comes from the division between the two disciplines seen in Hatch’s (1979) important paper in which she warns us that we need to be mindful of potential leaps in logic and warns us to ‘Apply with caution’ when we ‘apply research findings to classroom teaching’ (p. 123).

Nonetheless, even as late as the 1990s, one could question whether we are dealing with one discipline or two separate ones. Scholars in each of the two disciplines continued to see themselves as separate from one another, and professional organizations reflected the distance with debates about the identity of professional organizations (see Gass, 2001). Approximately 20–25 years ago one witnessed the fields of language teaching and language learning go their separate ways with the development of journals (e.g., Studies in Second Language Acquisition, Second Language Research) and conferences (e.g., SLRF, European SLA [EuroSLA], PacSLRF) focusing on issues of SLA, as separate from language teaching. The argument that we put forward in this paper is that as SLA evolved into a unique area of inquiry seeking to understand how second languages are learned without an emphasis on how they are taught, the field itself also focused on its capacity to produce scientifically rigorous research and thereby achieve respect as an independent discipline. Part of the respect comes from an emphasis on the kind of research we do and the kind of analyses we undertake. Statistical trends and an emphasis on research methods and design are strong indications of disciplinary sophistication has also been echoed by others (Gass, 2001; Khany & Tazik, 2019; Lazaraton, 2005; Loewen & Gass, 2009; Plonsky, 2013; Plonsky & Gass, 2011).

Having made the argument of two separate disciplines, we do not intend to suggest that they are unrelated. In fact, we would argue that the two disciplines feed one another and are relevant to one another. One only has to look at the burgeoning field of Instructed SLA (e.g., Loewen & Sato, 2017 as well as the new journal Instructed Second Language Acquisition, established in 2017 and published by Equinox) to understand the closeness of the two disciplines. What we do argue is that while the two are relevant to one another, there may be aspects of each that are not. For example, for many working within a generative SLA framework the main raison d'être for conducting SLA research is to understand the nature of language and the human capacity for language. One further indication of the separation of the two is the fact that in early days of SLA research, it was common to have an ‘implications for teaching’ section at the end of a paper. In today’s research climate, this is not so.

Khany and Tazik (2019) conducted an analysis of the content of ten journals over a nearly 30-year period (1986–2015). Their database consisted of 4,079 studies. Table 1 gives an indication of the research traditions that were covered, broken down by ten-year periods.

The first observation reflects the decrease in nonempirical studies over this time period. This reflects earlier comments whereby we showed an increase over time of articles focusing on L2 data analysis as opposed to articles focused on language pedagogy. A second observation reflects the
increased diversity in research types: qualitative research and mixed-methods research have increased in the past decade and there is a balance between quantitative and qualitative research.

More germane to our argument of increased sophistication is the classification of types of statistics. Khany and Tazik (2019) classified articles in their database using the Goodwin and Goodwin (1985) classification scheme of basic, intermediate, advanced. Goodwin and Goodwin’s initial classification considered articles only in one journal (American Educational Research Journal) over a five-year period (1979–1983) in an effort to determine the degree to which graduate students were prepared to read the educational literature. A brief summary of their categorization, as utilized by Khany and Tazik, is given below:

- **Basic**
  - Descriptive
  - Pearson Correlation
  - Chi-Square
  - t-test
  - One-way ANOVA

- **Intermediate**
  - Univariate
  - Nonparametric

- **Advanced**
  - Multivariate analyses

In Table 2 we present the results of Khany and Tazik’s analysis.

As is clear, the percentage of articles at both the intermediate and advanced levels has increased over the 30-year time period considered with a slight decrease in the percentage of articles using only basic statistics.

In another attempt to determine statistical sophistication, Gass (2009) presented the results in Figure 1 representing trends in SLA articles. The data are based on 1,411 articles appearing in four journals whose focus is on L2 learning (Applied Linguistics, Language Learning, Second Language Research, Studies in Second Language Acquisition). In her study, four categories were used to
determine statistical sophistication: articles with no statistics; articles with descriptive statistics only; articles with single inferential statistics; and articles with multiple inferential statistics. Assuming statistical use and sophistication as a measure of progress, we see that somewhere around the early 1990s there is a shift where single and multiple inferential statistics are on the increase and articles with no statistics and/or descriptive statistics only are on the decline.

The results from Plonsky (2014) also support this position. He examined the type and frequency of statistical analyses from the 1990s into the 2000s. Using a sample of 606 studies published in Language Learning and Studies in Second Language Acquisition, Plonsky found that the variety of statistical analyses per study increased from the 1990s to the 2000s; the study also revealed greater use of a number of more sophisticated analyses such as MANOVA and structural equation modeling.

The early 1990s represent an important time for SLA inasmuch as there is evidence that the field at about this time shows signs of becoming an independent field not tied with language teaching. But perhaps more important is that it was also a time when SLA was beginning to become normalized in the sense that there was an independent body of accepted knowledge. Evidence for this assertion comes from two main sources. First, at about this time the first textbooks appeared (Ellis, 1994; Gass & Selinker, 1994; Larsen-Freeman & Long, 1990). Second, a major conference dedicated to the discipline of SLA (Second Language Research Forum) took on a national scope. When SLRF began in 1977, it took place at one of two Los Angeles universities (UCLA and the University of Southern California, both of which had a group of L2 researchers) even though participants came from outside of the immediate area. In 1988, the University of Hawai‘i hosted SLRF, but importantly for our purposes, after 1991, the conference truly became a national (if not international) conference rather than a local California one, with conferences being held over the next decade in Michigan, Oregon, Pennsylvania, New York, Arizona, and Canada. (see https://sites.google.com/site/secondlanguageresearchforum/home for a list of all SLRF conference sites). A similar history can be presented showing the early 1990s as a period of growth for SLA by looking at SLRF-type/SLA conferences around the world. PacSLRF held its first conference in 1992 and, although not as regular, has held seven conferences in the Pacific area. Similarly, in a different part of the world EuroSLA (European SLA) began hosting an annual conference in 1991 (Salzburg, Austria).

In sum, we have shown that the field of SLA has gone through a period of time when it was linked (probably inextricably) to language teaching, to a time when there was an identity crisis as to its focus, to a period when it showed signs (conferences, journals, books) of an independent discipline. This development is paralleled by more advanced statistical approaches and measurements as an indication of disciplinary rigor. In today’s world, we are concerned with issues that all sciences are faced with, namely how our discipline can create results that are robust, credible, and reproducible. In fact, on National Public Radio (NPR, a U.S.-based news organization) on May 20, 2018, the following interview (www.npr.org/2018/05/20/612747674/magdalena-skipper-is-named-new-chief-of-nature) took
place between Lulu Garcia-Navarro from NPR and Magdalena Skipper, the new editor-in-chief of the science journal *Nature*. Part of this interview is reproduced below.

Skipper: Science itself changes, and so the publishing of science needs to change and evolve with it.

Garcia-Navarro: In what way?

Skipper: So, for example, right now scientists themselves are spending much more time and thinking more carefully about how robust, credible and reproducible the science is. And *Nature*...has actually led the way in helping scientists very clearly state how research was done, how the data are being generated and...

Garcia-Navarro: Transparency.

Skipper: Transparency, indeed. Transparency to the process, transparency of the results that leads scientists to make discoveries and come to their conclusions.

These are precisely the issues that our discipline is grappling with. The interview refers to *science* and *scientists*. Some would argue that this is an appropriate term for SLA research. To wit, the journal *Studies in Second Language Acquisition* has a submission category called 'State of the Science' which replaces the more common and earlier category 'State of the Art'.

We turn next to a discussion of where we are now in terms of our trajectory toward becoming a more methodologically mature and established discipline. In the final section, we address issues of the future and what indications there are that we are concerned with the issues mentioned by Skipper of robustness, credibility, and transparency.

**The Present**

We begin this section by noting, as was indicated earlier, that this paper was first presented at a conference devoted to research methodology in the field of SLA. The very existence of the Montpelier conference speaks to the current level of interest in research methods in the field.\(^6\)

In 2011, Plonsky and Gass commented, ‘Although scarcely mentioned in SLA, a robust line of research exists among other social sciences around the construct of study quality’ (p. 326), which Plonsky (2014) defined as (a) adherence to standards of contextually appropriate, methodological rigor in research practices, and (b) transparent and complete reporting of such practices’ (p. 451). The situation has changed some since 2011, and just two years later, Plonsky (2013) stated, ‘There is no controversy over the necessity of rigorous quantitative methods to advance the field of SLA’ (p. 656). In that same year, Byrnes observed that the field was undergoing a methodological turn, as we discussed earlier.


In addition to general research methods books, we also have discipline-specific books focused solely on quantitative methods, such as Loerts, Lowie, and Seton’s (2019) *Essential statistics for applied linguistics: Using R or JASP*. There is also Plonsky’s (2015a) edited volume *Advancing quantitative methods in second language research*, which contains chapters on various inferential statistical analyses that go beyond the more commonly used ones in our field. Another book that initially came out in 2010, but has a new edition, is Larson-Hall’s (2015) *A guide to doing statistics in second language research*.\(^2\)
research using SPSS and R. The first edition focused solely on SPSS, but the second edition now includes instructions for using R as well.


In addition to these books, there are several special issues of journals, such as Language Learning’s (2015) volume focusing specifically on quantitative analysis, as well as Applied Linguistics’ (2016) volume on innovation in research methods in applied linguistics. And, finally, there are numerous journal articles on the topic of quantitative research methods. Several of the more recent and prominent ones include Lindstromberg (2016) Inferential statistics in language teaching research: A review and ways forward, Norouzian and Plonsky (2018) Eta- and partial eta-squared in L2 research: A cautionary review and guide to more appropriate usage, Mizumoto and Plonsky (2016) R as a lingua franca: Advantages of using R for quantitative research in applied linguistics, and a favorite of ours, Al-Hoorie and Vitta’s (2019) The seven sins of L2 research: A review of 30 journals’ statistical quality and their CiteScore, SJR, SNIP, JCR impact factors. Finally, Larson-Hall’s (2017) Moving beyond the bar plot and the line graph to create informative and attractive graphics is a challenge to all of us to improve our data visualization techniques.

While many of these articles provide guiding principles for conducting and reporting quantitative analyses, there are an increasing number of journals that provide guidelines that authors must follow in order to submit manuscripts. For example, Studies in Second Language Acquisition has a section on quantitative research which states, among other things, a requirement for ‘completeness and transparency in quantitative data reporting practices.’ Language Learning has a page of submission requirements which specifically refers prospective authors to Norris, Plonsky, Ross, and Schoonen’s (2015) Guidelines for reporting quantitative methods and results in primary research. TESOL Quarterly (a journal which is focused on research that is relevant to or even feeds into teaching) very recently updated its author submission guidelines on the journal’s website with reference in particular to standards for quantitative, qualitative, and mixed-method research (see also Mahboob et al., 2016). Prior to this update, Brown (1991, 1992) served as important guidelines for researchers in both language teaching and language learning.

In addition to the proliferation of publications on quantitative research methods, another sign of increased interest is both seen in and draws on the growth of meta-analysis in the field. Plonsky’s bibliography of meta-analyses and papers related to meta-analysis (https://lukeplonsky.wordpress.com/bibliographies/meta-analysis/) includes over 300 references nearly all of which were published/presented in the last 20 years. One major impetus for the proliferation of this type of research and the scrutiny it often applies to different substantive domains is Norris and Ortega’s (2000) meta-analysis on the effectiveness of L2 instruction. Not only did the authors introduce much of the field to the potential of meta-analysis as a set of techniques for synthesizing results across a large body of empirical evidence, their study also called for ‘more rigorous practices for experimental and quasi-experimental designs’ (p. 502), thereby calling researchers’ attention to a number of methodological issues.

In the years since the publication of Norris and Ortega (2000), meta-analytic techniques have been applied widely as a means to both describe and evaluate L2 research methods. Research of this type, often referred to as ‘methodological synthesis’, generally falls into one of two categories. The first focuses on the research and reporting practices as observed in a particular substantive domain. Examples of this type include methodological syntheses of research on L2 written feedback (Liu & Brown, 2015), learner corpora (Paquot & Plonsky, 2017), task-based language production (Plonsky & Kim, 2016), and computer-based interaction (Ziegler, 2016).

The other category of methodological syntheses looks across domains to examine the use of one or more individual research techniques. Recent examples in this vein include Hashemi and Babaii’s (2013) review of mixed-methods research; Marsden, Morgan-Short, Thompson, and Abbugaber’s (2018) review of replication research; Plonsky, Marsden, Crowther, Gass, and Spinner’s (in press) synthesis of the use...
of judgment tasks; and Plonsky and Ghanbar’s (2018) synthesis of the use (and misuse) of multiple regression in L2 research. In both types, the goal remains largely the same: to describe and evaluate research and reporting practices as a means to provide empirically grounded recommendations for future studies. Based on their findings, for example, Marsden et al. (2018) outline a set of 16 suggestions for improving the culture and practice of replication research in applied linguistics.

Guidance from methodological synthesis on its own is, of course, not sufficient. In order for the field to conduct and report quantitative studies appropriately, researchers need to be trained. Several studies have looked at researchers’ statistical training and knowledge. For example, Loewen et al. (2014) found that only 14% of doctoral students and 30% of professors felt that their statistical training was adequate. In terms of statistical literacy, which Gonulal (2016) defines as ‘the ability to (a) understand basic statistical terminology, (b) use statistic techniques appropriately, and (c) interpret statistical analyses’ (p. 18), it is clear that there is a range of proficiency levels in the field. However, statistics courses offered by SLA graduate programs do lead to students reporting greater feelings of statistical self-efficacy (Gonulal, Loewen, & Plonsky, 2017). In a review of 14 graduate program requirements, it was found that 40% required one quantitative research methods course, while 21% required three courses. One program did not require any quantitative methods course.

In addition to classroom training, there are numerous workshops that researchers in the field can avail themselves of. In recent years, the American Association for Applied Linguistics (AAAL) and SLRF have held pre-conference workshops on both qualitative and quantitative methodology. Furthermore, we are aware of three different summer schools, workshops, and ‘bootcamps’ that have recently taken place, all devoted to research methods in applied linguistics.

Thus, it seems like it is a good time for someone to be interested in quantitative methodology in applied linguistics and SLA research. There are advances currently underway in numerous directions.

The Future

We are encouraged by the progress that the field has made toward a greater awareness of and emphasis on the importance of sound research methods. However, great strides are still needed in order for the field to maximize its potential to inform our understanding of L2 learning, teaching, use, and assessment. As an initial step, we would like to propose a definition of ‘study quality’ that expands upon a traditional understanding of this term as research that is carried out according to standards of rigor (e.g., appropriate design, valid instrumentation, justifiable analyses). This definition is useful, but it is incomplete. Specifically, we would add that high quality research also possesses the following three additional characteristics.

First, it is primarily concerned with estimating the magnitude of effects. Rather than seeking the mere presence or absence of an effect or relationship, high quality research provides an indication of the extent that such an effect or relationship is observed (Norris, 2015; Plonsky, 2015b).

Second, high quality research is transparent. Transparency can take many forms, but here we refer in particular to researcher behaviors and practices following the completion of a study such as (a) honest and thorough (as opposed to selective) reporting of results, and (b) making materials and data available to other researchers for inspection and/or re-analysis (for a discussion on data sharing, see Plonsky, Egbert, & LaFlair, 2015).

The third characteristic we would add to our definition is that of reproducibility. In order to consider a particular study to be of high quality, its findings should be able to be reproduced. Part of ensuring quality in this case involves the previous notion of transparency; any lack of critical procedural details therefore poses a threat. Reproducibility, however, also implies a degree of stability and accuracy such that a study’s findings could reasonably be expected to replicate given a comparable sample.

Drawing on this expanded definition, and building on the developments that we have outlined thus far, we would like to describe what we see as five directions for the field’s continued use and advancement of quantitative methods.
Open science
The place and value of open science practices is closely linked with the notion of transparency described above (see Marsden & Plonsky, 2018). However, open science is also manifested in many different ways. For example, we envision and hope to see the field move toward more thorough reporting of quantitative data. Doing so not only increases the meta-analyzability of primary studies (Larson-Hall & Plonsky, 2015), it also allows consumers to better understand, interpret, and evaluate the studies they read. As stated by the American Statistical Association (Wasserstein & Lazar, 2016), ‘Proper inference requires full reporting and transparency’ (p. 131).

Practicing open science goes well beyond thorough reporting of data. It is critical that we also foster a culture of collaboration in which researchers share and make available their materials on repositories such as IRIS (https://www.iris-database.org/iris/app/home/index;jsessionid=08652A75E8B0264C36BD0DE5C28E1E56). Such efforts increase field-wide efficiency, facilitate replicability, and support researcher training, among other benefits (Marsden, Mackey, & Plonsky, 2016).

Replication
A second direction we hope to see the field taking is toward increased replication. As revealed very clearly in Marsden et al.’s (2018) synthesis, a very small portion of L2 research to date has had an explicitly declared interest in replicating previous findings. Part of the lack of such studies may be due to a number of misconceptions about the place and value of replication research. Anecdotally, it is our experience that many scholars continue to devalue replication studies due to their lack of novelty (cf. discussion in Porte & McManus, 2019, p. 4). Others hold the mistaken view of replication research primarily as a means to challenge or find fault in an existing study. This kind of self-correction is certainly part of the scientific process, but it is rarely the main goal. Rather, replication studies are more appropriately viewed as a means to (a) refine results and test their generalizability (e.g., for different populations, target structures, settings; see Plonsky, 2012); (b) arrive at more stable conclusions which are often difficult to obtain based on the small samples typical of L2 research; and (c) teach/learn how to conduct a study (see Porte & McManus, 2019, for a practical guide to conducting replication research).

In addition to these meaningful contributions, we would also note that replication research can entail a number of other benefits for scholars. For example, Marsden et al. (2018) found that replications tend to be cited approximately six times more often than nonreplication studies. Furthermore, in recognition of the value of replication research, the tenure and promotion guidelines of American Association for Applied Linguistics (AAAL, n.d.) now urge annual review committees to consider ‘high quality replication studies…on par with nonreplication-oriented studies.’ We very much hope and expect that these rationales will encourage an increase in replication research in the field.

Improving and advancing statistical analyses
We have generally seen two types of advances taking place in the field’s quantitative analyses: improved use of familiar techniques such as multiple regression (Plonsky & Oswald, 2017) and the introduction of novel techniques such as bootstrapping, and Bayesian data analysis (Larson-Hall & Herrington, 2010; Norouzian, de Miranda, & Plonsky, 2018). As we continue on the path toward methodological improvement, we anticipate further growth in both of these two areas.

One potentially fruitful avenue for introducing new techniques involves opening ourselves to the work of colleagues in other departments such as psychology or economics, both of which have long-standing traditions of methodological expertise and sophistication. To be clear, we are not arguing for technification for the sake of technification. In some cases, however, more advanced statistical modeling can lead to insights that cannot be accessed using simpler or more traditional analyses (see Brown, 2015).

Methodological expertise, training, and research
In order for the improvements in statistical analyses and other research practices described here to take place, more education is certainly needed (Gonulal, forthcoming). Training of this type will most likely
and most often occur in formal settings such as graduate coursework and organized workshops. However, the curricular changes needed to advance the field in this way will require foresight and planning on the part of graduate faculties. Given the potential tension between the desire to include such classes and associated staffing challenges, we would encourage graduate programs to, once again, allow if not encourage students to take classes in other departments where greater expertise can be found. Students can then share what they have learned with their home departments and in the form of conference papers and published articles that demonstrate how novel and perhaps more nuanced techniques might be applied to L2 research.

Wherever methodological training takes place, we encourage students as well as more established scholars to develop methodological specialties, both for their own work as well as for the betterment of the field. Of course, this assumes that critical thinking skills and the ability to critically read the literature form the basis of all graduate education. These skills are precursors to methodological training. We envision a future in which L2 researchers identify not only as having expertise in substantive domains but as experts in one or more methodological techniques as well.

**Expanding the role of journals and learned societies**

Many of the changes and advances we have discussed thus far require action on the part of individual researchers. Improvements can—and, we would argue, should—also occur in a kind of ‘top-down’ fashion. One example of top-down reform can be observed in the requirement that many journals have instated to report effect sizes. Following Ellis’s (2000) editorial announcing this change in policy, a move since taken up by virtually all major applied linguistics journals, the reporting of effect sizes has greatly increased (see Plonsky & Oswald, 2014; Rining, Hu, & Xiong, 2019). But there is certainly more that can be done.

Journals can continue to refine their guidelines and to insist on methodological quality. We believe that researchers will rise to the occasion. Journals can also dedicate space for the discussion of methodological issues.

We would also encourage a greater role from learned societies such as AAAL and EuroSLA. There are currently no field-specific standards that apply to all of applied linguistics. Norris et al.’s (2015) guidelines are useful and represent a strong start, but they pertain only to quantitative L2 research.

Professional organizations might also consider weighing in on curricular issues. It is not unreasonable to expect that AAAL could provide guidance to graduate programs with respect to what might be expected for methodological training. Here, too, we can turn to other more established disciplines that might serve as models. The Society for Industrial and Organizational Psychology (Division 14 of the American Psychological Association), for instance, recently published a set of guidelines for graduate education which include a clear description of methodological and statistical topics to be covered (Gibson, Payne, Morgan, & Allen, 2018).

**Conclusion**

We have shown in this paper that the empirical methods employed in L2 research have undoubtedly advanced since the field’s inception. The increased attention to methodological issues observed in recent years should inspire greater confidence in our understanding of the constructs and relationships of interest to SLA. However, as demonstrated by the growing body of methodological syntheses, improvements are needed across a wide range of substantive and methodological domains. And the stakes are high. By dismissing or overlooking weak designs and inappropriate analyses, for example, we run the risk of committing massive inefficiencies, misinforming practitioners, and losing credibility among our peers in other social sciences. At the same time, we have good reason to be optimistic. As we have laid out in the final section of this paper, a future of continued methodological reform and improvement in SLA is within reach.
Notes

1 A U.S. agency responsible for health, biomedical, and public health research.

2 Foreign Language Annals, System, Applied Psycholinguistics, Applied Linguistics, Language Learning, Modern Language Journal, Studies in Second Language Acquisition, TESOL Quarterly Language Teaching, Language Testing. Unfortunately, their mix of journals includes journals reflecting both language teaching and language learning and one dealing with language testing. There is no way to separate which data come from which journal. Nonetheless, we feel that these data give us an indication of the role of article types and statistical usage in all of the fields covered by these ten journals.

3 The authors did not define their categories. We assume that nonempirical studies include, for example, theoretical treatises. Quantitative and mixed-methods studies include some sort of statistical analysis, with the latter also including qualitative descriptions of data. Qualitative studies only include qualitative descriptions with no statistical analyses.

4 Clearly, we cannot become too complacent. Lindstromberg (2016), in a survey of the first 19 volumes of Language Teaching Research (through 2015), reminds us that there is still a limited range of statistics used, small sample sizes, and frequently there is no indication of effect sizes.

5 Prior to this time, courses in SLA (when they existed) relied on readings and did not have the benefit of received knowledge in one location, as is provided by textbooks.

6 The current interest in methodology is further evidenced by another conference with similar themes: Georgetown University Round Table in 2012, which was titled Measured Language: Quantitative Approaches to Acquisition, Assessment, Processing, and Variation. A second important innovation is the recent addition of a research methods strand at the annual AAAL meeting.

7 The sins include (1) not reporting reliability, (2) not discussing validity, (3) making inferences from descriptive statistics, (4) incomplete reporting, including nonsignificant results, (5) not reporting effect sizes, (6) not adjusting for multiple comparisons, and (7) not reporting assumption checks.

References


Susan Gass is Professor in the School of Foreign Languages at Southeast University in Nanjing, China and University Distinguished Professor in the Second Language Studies Program at Michigan State University. She has published more widely in the field of SLA, including books on research methods (Data elicitation for second and foreign language research and Second language research: Methodology and design), both published by Routledge and co-authored with Alison Mackey. Many of her books/articles have been translated into Arabic, Chinese, Korean, and Russian. She has lectured in many parts of the globe and is the recipient of numerous local, national, and international awards. She has held leadership positions at her university and has also served as the President of the AAAL and the International Association of Applied Linguistics. She is the current editor of Studies in Second Language Acquisition.

Shawn Loewen (Ph.D., University of Auckland) is Professor at Michigan State University in the MATESOL and Second Language Studies programs. He also serves as director of the Second Language Studies program. His research interests include instructed SLA, L2 interaction, and quantitative research methodology. In particular, his current interests include the link between SLA research and researchers, on the one hand, and second language teachers and pedagogy on the other. In addition to publishing in leading SLA journals, he has co-authored two books, Key concepts in second language acquisition (with Reinders, 2009), and An A-Z of applied linguistics research methods (with Plonsky, 2016). His sole authored book, Introduction to instructed second language acquisition, appeared in 2015, and the co-edited The Routledge handbook of instructed second language acquisition (with Sato) was published in 2017. In 2018, he became the associate editor of Modern Language Journal.
Luke Plonsky is Associate Professor of Applied Linguistics at Northern Arizona University, where he teaches courses in second language acquisition and research methods. Recent and forthcoming publications in these and other areas can be found in *Applied Linguistics*, *Language Learning*, and *Modern Language Journal*, among many other journals and volumes. He has also written and edited several books. Luke is Senior Associate Editor of *Studies in Second Language Acquisition*, Managing Editor of *Foreign Language Annals*, Co-Editor of de Gruyter Mouton’s Series on Language Acquisition, and Co-Director of the IRIS repository for instruments in language learning and teaching (iris-database.org). Luke held previous faculty appointments at Georgetown University and University College London. He has also taught in Japan, The Netherlands, Poland, Puerto Rico, and Spain. Luke received his Ph.D. in Second Language Studies from Michigan State University.