Reliability and Validity in SLA Research
Workshop Programme
11-12 June 2015

LOCATION:
Facultat de Filologia, Universitat de Barcelona
SALA GABRIEL OLIVER (Josep Carner Building, basement)
C/ Aribau, 2
08007 Barcelona

With the support of:

EuroSLA Workshop Grant Programme
PhD Programme in Applied Linguistics (UB)
MA Programme in Applied Linguistics and Language Acquisition in Multilingual Contexts (UB)
Departament de Filologia Anglesa i Alemanya (UB)
Facultat de Filologia (UB)

Thursday, 11th JUNE 2015

9:30-10:00	**Workshop Registration** (Sala Gabriel Oliver – Josep Carner Building, basement)
10:00-11:00	**Keynote speaker: Rob Schoonen** (Amsterdam Center for Language and Communication - University of Amsterdam, The Netherlands) "**Generalizability, where theory and methodology meet**" (see the last page for the full abstract)
11:00-11:30	**Coffee Break**
11:30-12:00	**Luke Plonsky** (Northern Arizona University, USA) **"Sample Demographics and Generalizability in SLA: A Systematic Review"** Using a sample of 604 primary studies, I systematically review the demographics in SLA across educational levels, first languages, age groups, and so forth. The results show oversampling of young adults and speakers of English as an L1 or L2, among other populations. Suggestions for improving external validity are discussed.
12:00-12:30	**Lidia Montero** (Universitat de Barcelona, Spain) **"Double analysis approach in L2 communication strategies effectiveness"** This paper offers an explanation of the analysis methods implemented in a study on L2 communication strategies effectiveness. The operationalisation of this context- dependent construct presented a challenge faced by means of the triangulation of two different analysis approaches that not only complement each other but also show a positive correlation.

12:30-13:00	**Sofía Moratinos** (Universitat de les Illes Balears, Spain) **"The role of the language learning context and how it affects motivation"** As motivation plays a key role in determining success in language learning, it is worth examining how particular language learning contexts- e.g. formal classroom instruction, CLIL and study abroad, that students can experience during their education- may affect motivation. This is the idea behind the Language Learning Context and Motivation Questionnaire presented in this paper.
13:00-15:00	**LUNCH**
15:00-15:30	**Dakota Thomas, Jennifer Ament** and **Eloi Puig-Mayenco** (Universitat Pompeu Fabra, Spain) **"Collecting quantitative data: the web-based experience"** The talk aims at presenting three studies. The commonality of the three studies is the use of web-based tools to collect quantitative data. By presenting the methodology and remarks of each study, the intention is to discuss the advantages and limitations of these softwares when collecting SLA data.
15:30-16:00	**Alexandre Nobajas** (Keele University, UK) **"Is Twitter a valid data source to study linguistics? The case of Catalonia"** Twitter is increasingly being used to gather linguistic and sociolinguistic data, but it is necessary to assess it in order to use it for scientific research. Topics such as data validity, user demographics, language detection systems or geocoding are discussed and explained in order to evaluate the Twitter as a data source.
16:00-16:30	**Marina Ruiz** (Universitat de Barcelona, Spain) **"Photo-Initiated Statuses on Facebook by Bilingual Mothers: Data Collection and Reporting"** This study examines Japanese-English bilinguals and their interactional processes in online conversation on social networking sites (SNS). This presentation will detail the methodology used to obtain this particular data set, and the justification of the data sets that were used for reporting.
16:30-17:00	**Coffee Break**
17:00-18:00	**Keynote speaker: Martin Lamb** (School of Education – University of Leeds, UK) "**Researcher effects in qualitative inquiry**" (see the last page for the full abstract)

Friday, 12th JUNE 2015

9:30-10:00	**Elena Markina** and **Joan Castellví** (Universitat de Barcelona, Spain) **"Measuring learner's accuracy and complexity in Russian: constraints and solutions"** The goal of this paper is to analyse some of the issues related to applying complexity and accuracy measures to Russian on basis of the data collected in a series of written and oral tasks with 18 Catalan learners of Russian with low level of proficiency.
10:00-10:30	**Ghulam Abbas Khushik** and **Ari Huhta** (University of Jyvaskyla, Finland) **"Exploring the findings of the Facets, Syntactic Complexity Analyzer and Coh-Metrix in writings of Finnish learners of English as a second language"** This study reports on the understanding of measures of multifaceted Rasch program Facets and the different findings for same variables in Coh-Metrix and the Second Language (L2) Syntactic Complexity Analyzer applications which are not well explored. The study aims to investigate the most reliable Syntactic Complexity metrics.
10:30-11:00	**Tim Diaubalick** (University of Wuppertal, Germany) **"Individual Differences – The Problem of Forming Groups in SLA Aspect Studies"** This PhD-project investigates the SLA of Spanish Aspect, being based on particular findings of a sample of German students who considered adverbials only. To discover the cause of this behaviour, various individual variables (L1, instruction,…) are taken into account. Feasibility will be improved by forming as few groups as possible.
11:00-11:30	**Coffee Break**
11:30-12:00	**Ferran Gesa** and **Imma Miralpeix** (Universitat de Barcelona, Spain) **"Examining Productive Vocabularies by means of *V_Size*: Some Preliminary Considerations"** This paper uses a new instrument (*V_Size*) to measure productive vocabulary size. The study evaluates its discriminatory power between learners and native speakers or the appropriateness of baseline data. Results show that *V_Size* discriminates between different proficiency levels and that analyses can be more accurate when evaluating narratives with NSs' data as the baseline.
12:00-12:30	**Elke Peters** and **Maribel Montero Perez** (KU Leuven, Belgium) **"Vocabulary acquisition through audio(-visual) input: how to measure learning gains?"** This study discusses the methodological challenges in research into vocabulary acquisition through audio(-visual) input. Specifically, we focus on the design, the selection of target items, the measurement of vocabulary gains, and the data collection procedures. Our findings seem to point to a trade-off between ecological validity and experimental control.
12:30-13:30	**Keynote speakers: Luke Plonsky** and **Deirdre Derrick** (Northern Arizona University, USA) "**Reliability Coefficients in L2 Research: A Meta-analysis**" (see the last page for the full abstract)
13:30-13:45	**CLOSING**

KEYNOTE SPEAKERS' ABSTRACTS

Rob Schoonen
(Amsterdam Center for Language and Communication - University of Amsterdam, The Netherlands)

"Generalizability, where theory and methodology meet"

One of the primary goals of scientific research in SLA is to gain insights that will help us to make decisions on, or predictions about, language learning. An underlying assumption is that the data we have collected, allow us to make valid decisions and predictions, that is, decisions and predictions that are generalizable to new cases, other measurements, and maybe new situations. In current L2 research, and applied linguistic research in general, little attention is given to the generalizability of measurements and research outcomes. As a research community, we run the risk of building interpretations on incidental outcomes. In the first part of the presentation a number of these 'generalizability' problems will be outlined in more detail, such as sampling of participants, sampling of measures and also interpretation of statistical outcomes. In the second part, I will focus on measurement and show that psychometric issues are in principal theoretical issues. Finally, a few possible ways to deal with these generalizability issues will be discussed. Answers to questions about generalizability of research findings are needed in order to be able to build a solid and generalizable knowledge base in our field of SLA.

Martin Lamb
(School of Education – University of Leeds, UK)

"Researcher effects in qualitative inquiry"

Experimenter/investigator/observer effects – and other variants on the original Hawthorne effect - are well-known in scientific research as potential threats to reliability and validity. In qualitative research they are not usually considered as threats since the reflexive researcher must always consider their own role in the research procedure, how they may be influencing the behaviour of participants and introducing bias into their analysis and interpretation of data. In this talk I will review these necessary precautions, but also consider the possibility that our qualitative studies have longer-term effects on our participants, perhaps years after the research has ended, and these may have ethical implications.

Luke Plonsky and Deirdre Derrick
(Northern Arizona University, USA)

"Reliability Coefficients in L2 Research: A Meta-analysis"

Concerns have been raised recently over the lack of reliability estimates found in reports of published L2 research. A potentially greater threat to internal validity is the actual reliability of data collection instruments being employed. Unfortunately, L2 researchers are left to interpret the performance of their instruments, if they do so, in relation to generic benchmarks. In response to these issues, this paper meta-analyzes reliability estimates (internal consistency, interrater, intrarater) as reported in the field. We also explore heterogeneity in observed estimates as a function of different coefficients and of study and instrument features suggested to moderate them. More concretely, we recorded 2,244 reliability coefficients found in 532 published studies of L2 research along with a number of features corresponding to the designs (e.g., sample size, proficiency) and instruments (number of items, test formats) employed. We also coded for the type of reliability index reported (e.g., alpha, KR20). The three types of reliability varied, with the lowest estimates found for internal consistency: median = .82. Interrater and intrarater estimates were substantially higher at .92 and .95, respectively. These overall estimates were also found to vary according to several study and instrument features such as learner proficiency (low = .79, intermediate = .84, advanced = .89) and target skill (e.g., writing = .88 vs. listening = .77). In addition to raising L2 researchers' awareness of the need to report reliability and other psychometric features, we use our results to inform and encourage interpretations of reliability relative to the larger field as well as to the substantive and methodological features particular to individual studies and subdomains.