THE LANGUAGE BACKGROUND QUESTIONNAIRES IN L2 RESEARCH: TEASING APART THE VARIABLES

Laura Sabourin, Jean-Christophe Leclerc, Myriam Lapierre, Michèle Burkholder and Christie Brien
University of Ottawa

1. Introduction

The psycholinguistic study of bilingualism and the process of becoming bilingual is a difficult undertaking for many reasons, not least due to the fact that bilinguals and second language (L2) learners are not homogenous groups. In short, the individuals that make up these groups differ in how, when, and how well they have acquired their two languages. There are many factors that impact how bilinguals and L2 learners acquire, process and represent their languages in their mind. In order to tease apart the complex variables at play, the field of psycholinguistics would greatly benefit from the availability of a comprehensive and widespread language assessment tool that can be used to adequately categorize participants. Our goal is to describe such a tool: a language background questionnaire (LBQ) through which researchers can gather age of L2 acquisition (AoA), manner of acquisition (MoA), and self-rated proficiency data, among other types of data.

In Section 2, we outline why our LBQ is needed and how it improves on other similar tools already in use. In Section 3, we then discuss (non-exhaustively) a number of important grouping variables currently used in bilingualism and L2 research, including manner of acquisition, a variable that has not been used as extensively as a grouping variable in psycholinguistic research, but appears to be crucial based on recent research (e.g., Sabourin, Leclerc, Burkholder, and Brien 2014; Pliatsikas and Marinis 2013). In Section 4, we introduce the components of our LBQ and report on how we have tested it for validity, reliability and efficiency. We conclude in Section 5 by summarizing our LBQ and discussing its strengths and how we are continuing to improve it.

2. Background

In psycholinguistic studies involving sequential L2 learners, it is common knowledge that there is a need to somehow group or select participants based on a set of variables deemed relevant for each particular study. Typical variables that are reported in classical as well as current research include such factors as AoA and proficiency at the time of testing. Other factors that often influence the outcome of language learning are factors such as motivation and aptitude, however these are less often reported in the psycholinguistic literature. Because these factors can vary considerably from participant to participant, this can result in a very heterogeneous group of L2 participants and, if such diverse participant background factors are not carefully controlled, this can have varying, and often confounding, effects on experimental results.

Actes du congrès annuel de l'Association canadienne de linguistique 2016.
Proceedings of the 2016 annual conference of the Canadian Linguistic Association.
© 2016 Laura Sabourin, Jean-Christophe Leclerc, Myriam Lapierre, Michèle Burkholder, Christie Brien

Even simultaneous bilinguals, those who acquired both languages from birth, can exemplify a diverse group of participants. For example, some may exhibit incomplete acquisition and/or attrition of one of their languages. Further, exposure to both languages during infancy is rarely balanced, and even when it is, many simultaneous bilinguals may still show effects of language dominance and unbalanced proficiency. Adding further to the complexity described here is the dynamic nature of bilingualism, in which any grouping factor chosen for a study generally changes over the course of the lifespan.

An important question then is how to efficiently and effectively measure language knowledge, and all of its dynamic features, in order to investigate which factors affect how multiple languages are represented and processed in the brain. In order to do this, we need a language background assessment tool that is detailed enough to obtain information on all the relevant variables that may affect language processing, yet that is also simple enough that it is appropriate even for those with a straightforward language background.

Many research labs currently use either a questionnaire that they have developed themselves for a very specific research question, or one of the few publicly accessible language background questionnaires. Some examples can be found online at http://www.nhlrc.ucla.edu/nhlrc/data/questionnaires. One striking characteristic to note about many questionnaires in this list is the specificity of language pairings and type of participants (e.g., child vs. adult) that they are aimed at. Other widely available questionnaires include a web-based "Language History Questionnaire" (LHQ; Li et al. 2014) and the LEAP-Q (Marian et al. 2007). Both of these published questionnaires are available in numerous languages and aim to solicit information on linguistic background and self-rated proficiency. The web-based LHQ offers downloadable questionnaires as well as web-design features that allow users to modify the information being collected. While this is a useful feature, different studies may not necessarily be comparable if researchers are collecting different data from participants. The LEAP-Q is a very extensive questionnaire that specifically asks about participants' current knowledge and exposure to the L2, resulting in a short and efficient questionnaire. However, it does not examine the dynamic nature of language use and proficiency, and thus does not provide a complete picture. In addition, these questionnaires are aimed more at L2 learners and late bilinguals and are consequently not as appropriate for classifying simultaneous bilinguals.

An additional issue is that it is sometimes difficult to directly compare grouping variables from different studies, as the same variables may not always be operationalized in the same way. Our LBQ provides the means for researchers to operationalize these variables in a way that fits their research goals, but crucially, also the means to be explicit in communicating how they chose to do so.

While our LBQ may also be viewed as just another tool adding to those already available, we strongly believe that it has great potential to be the optimal standardized tool that can be used by researchers regardless of research focus and language groups being tested. Our LBQ also fills many of the gaps found in other existing questionnaires. For example, it allows researchers to measure changing amounts and types of exposure across time in a relatively simple manner, allowing this to be more easily comparable between and within studies. Obtaining information regarding such factors as incomplete

acquisition, attrition, changing proficiency levels, fluctuating language dominance, and differing contexts and environments allows researchers to better control the heterogeneity of their participant groups, leading to cleaner data and stronger conclusions. Next, we discuss grouping variables currently investigated in bilingualism and L2 research.

3. Grouping variables

3.1 Age of Acquisition (AoA)

AoA of the L2 is likely the most studied variable in the field of L2 acquisition. It is generally viewed that the earlier the L2 is acquired the more successful the learner will be in achieving native-like competence (e.g., Birdsong 1992), and this has led some to argue for a critical or sensitive period of language acquisition (e.g. Lenneberg 1967; Johnson and Newport 1989). Most researchers therefore agree that bilinguals and L2 learners need to be categorized by AoA; however, the way in which AoA is precisely quantified can differ from one study to another. AoA may represent the age of arrival into the L2 community, the age of first exposure (Ao1E) to the L2, the age of immersion in the L2 (AoI), or the age of instruction in the L2. All of these ways of quantifying AoA are valid, and may even correspond to the same age for particular individuals; however, selecting one over the other may result in a different pattern of results.

For example, in the Ottawa-Carleton District School Board, all students start learning French in Kindergarten, and so all would have an AoA of approximately 4 yrs old, if AoA is defined as Ao1E; however, depending on the type of program the students are in, they may receive only 20 minutes a day (a Core French program) or 150 minutes a day (Early Immersion program). A subset of the Core French students may later participate in a Late French Immersion program, where their exposure to French is increased significantly. As such, if AoA is defined as AoI, this would result in students in the Early Immersion program having a AoA of 4 yrs old, and the students in the Late French Immersion program having an AoA of 9-12 yrs old. On the other hand, the students in the Core French program would never have been immersed (i.e. no AoA), and so would be classified as functional monolinguals, which generally corresponds to how they actually self-identify. Essentially, if AoA is defined as Ao1E, there would be one group of "early bilinguals" with an AoA of 4 yrs old; however, if AoA is defined as AoI, the same participants would be divided into three groups: functional monolinguals, early L2 learners, and late L2 learners. Indeed, operationalizing AoA as AoI was determined to be crucial in a study involving the processing of non-cognate translation equivalents (Sabourin, Brien and Burkholder 2014), where adults who had an early AoI (under 7 yrs old) showed masked translation priming, while the participants with a later AoI (over 7) did not. These results are evidence that an early AoI leads to a shared bilingual lexicon.

Not only is it essential to be explicit in defining how AoA is operationalized, but it is also crucial to do so in a way such that results can be easily compared across studies. Thus, while researchers can choose, for example, to use Ao1E as the main grouping variable, they should also provide greater detail regarding the other dimensions of AoA, such as if/when participants were immersed, if/when they arrived in an L2 community,

and how much exposure to the L1 they have received. By providing a more comprehensive picture of AoA, a variable of considerable complexity, studies would be better equipped to address the conflicting evidence for critical/sensitive period effects.

3.2 Proficiency

Perhaps the second-most studied factor influencing bilingualism is proficiency. Like AoA, however, the mastery that any participant has over his or her languages is complex to quantify, due to the fact that participants may be stronger in certain skill areas than others (e.g., reading vs. speaking), and that it is difficult to find objective measures that are sensitive enough to capture these asymmetries. Indeed, there does not seem to be any good *quick* measure of overall proficiency that has been standardized, and this has led researchers to use proxy measures. These measures may only be testing at the phonological, lexical or syntactic level, or they may only test either expressive or receptive abilities. This requires the researcher to assume that their chosen proxy measure is a good overall indicator of proficiency. Even measures that are argued to provide a relatively global evaluation of proficiency, such as cloze tasks (which test competence at the morphosyntactic, lexical and discourse levels), have limitations in terms of the skills that they test (written but not oral), and the type of bilingual population for which they are appropriate (see Tremblay 2011 for further discussion).

In addition, it can also be quite difficult to compare proficiency test scores across participant groups and across languages. For example, if a sequential L2 learner of French scores higher on a French cloze task than a native speaker of French, does that mean that the L2 learner is more proficient, or simply that they learned standard French morphosyntax in a more explicit manner? If a simultaneous bilingual scores higher on a French cloze task than on an English one, does that mean that they are actually more proficient in French, or simply that the task is easier in French? Indeed, even if the same proficiency test is translated into both languages, the difficulty level is not necessarily matched. For example, what is being measured if the underlying syntactic structure of a particular "sentence" is different in the two languages, such as if object raising is involved in one language but not the other? If "objective" proficiency scores are not easily comparable, then it is difficult to match proficiency categories in any given study.

In contrast to these objective measures, many L2 studies use participants' self-ratings to assess language proficiency. While these are much more subjective and, as such, seem difficult to compare across participants, they have been shown to correlate significantly with more objective measures, such as cloze task scores (e.g., Sabourin, Brien and Burkholder 2014). Further, asking participants for explicit self-ratings has the benefits of being very quick, and of providing an opportunity to probe specific skills areas, such as oral comprehension, oral production, writing proficiency, reading proficiency and pronunciation. They also allow the researcher to enquire about participants' proficiency during different time periods (e.g current general proficiency versus highest attained proficiency), allowing researchers to also verify attrition. Though many studies ask only for a "global" proficiency self-rating, a more detailed view of language proficiency can also be obtained if the right questions are asked.

3.3 Dominance

One of the more dynamic factors that affect L2 and bilingual processing is language dominance (Treffers-Daller 2016). On the surface, this concept appears relatively simple: a bilingual's general preference for one language over the other; however, digging deeper, dominance may be due to a range of different sub-factors, such as unequal Proficiency levels, or higher frequency of use in certain contexts, such as in school, in the home, or in the workplace. Further complicating this is the fluid changes of environments over the lifespan. For example, a French-English bilingual who is immersed in an English-dominant university may later be immersed in a French-dominant workplace after graduation. One conceivable consequence of this is that the results of any assessment taken after a recent change in environments could be inaccurate if the full context is not taken into account. It is especially important to assess language dominance at the time of testing for participants who may have an AoA like that of one participant group, yet have a proficiency level similar to that of another group.

In order to properly assess language dominance, we suggest here that it is important to gather information regarding L1 and L2 usage at different stages of life (e.g., infancy, primary school, secondary school, etc.) and in different environments (e.g., school, home, work) in order to determine the relative use of each language, as well as how, why, and when this has changed over time.

3.4 Manner of Acquisition (MoA)

The fourth variable to which we will devote some attention is MoA, which has not been used as extensively as a grouping variable in psycholinguistic research, but does appear to be a key determiner of L2 processing based on recent findings (e.g., Sabourin, Leclerc, Burkholder and Brien 2014; Dussias and Sagarra 2007; Pliatsikas and Marinis 2013). MoA has been defined in various ways, with some authors focusing on the length of exposure (Dussias and Sagarra 2007) as opposed to the type of input (e.g., Sabourin, Leclerc, Burkholder and Brien 2014; Pliatsikas and Marinis 2013). We propose here to define MoA as whether a speaker's L2 was acquired in a naturalistic environment (such as in the home or in the playground), or whether the L2 was acquired in a more formal and explicit environment (such as in the classroom). For example, in the Ottawa-Gatineau region, bilinguals whose L1 is French are typically immersed in their L2 English in a relatively naturalistic environment, as they are constantly exposed to the more dominant language of the region in which they live, work, and play. Conversely, bilinguals whose L1 is English are typically immersed in their L2 French in a more formal setting, such as in the "French Immersion" classroom. As such, these two quite different learning environments may conceivably have an impact on different bilingual groups. For example, Sabourin, Leclerc, Burkholder, and Brien (2014), conducted a follow-up study to Sabourin, Brien, and Burkholder (2014), and found evidence that a naturalistic MoA was more important than an early AoA in promoting a shared bilingual lexicon.

Like dominance, we suggest that, in order to properly assess MoA, it is necessary to obtain information regarding L2 usage in different environments, and to compare exposure in those environments that are more naturalistic in nature to those that are more formal in nature. It is also crucial to find a way to operationalize MoA that can be applied in a comparative way to different bilingual communities in order to ensure that studies can effectively investigate its contribution to bilingual processing and representation.

3.5 Summary

The above section has demonstrated that there are multiple factors that are relevant to bilingual and L2 research, factors which are complex to define and measure. To further complicate things, many of these variables are often confounded with one another; for example, bilinguals who began acquiring their L2 at a young age typically achieve higher proficiency, and have benefited from a more naturalistic MoA. Thus there is a need for a dynamic and sensitive language assessment tool which enables these variables to be teased apart and evaluated in order to be used more effectively and appropriately as grouping variables in bilingual research. The following section describes such a tool that we have developed and provides an objective assessment of its validity and effectiveness.

4. Our LBQ

4.1 Components

The LBQ that we have developed[1] consists of two parts: a two-page Short Version, which provides a general overview of the participant's biographical and language background; and a five-page Extended Version, which provides complementary information that allows for a more fine-grained classification of participants with respect to those critical factors examined in Section 3. The Short Version takes an average of 8.5 minutes for the participant to complete (N = 66, SD = 3.55), and the Extended version takes an additional 19.5 minutes (N = 67, SD = 7.38). The data from both parts can be easily compiled and assessed by the researcher or assistants in just a few minutes using available templates[2].

The Short Version has two primary functions. First, it collects key language data that can be used to classify all participants with respect to AoA, (operationalized as either Ao1E or AoI), and provides a coarse-grained classification of participants based on proficiency and language dominance. As such, in studies where only such a coarse-grained classification of participants is required, the Short Version can stand alone as the only LBQ used. It also is able to identify whether participants have undergone attrition in any of their languages as it asks them to evaluate their current overall proficiency as well as the highest level of proficiency ever achieved for each of their languages. Second, for those studies requiring a more fine-grained classification of participants, the Short Version provides sufficient information to clearly classify participants as 1) functional

[1] Our LBQs can be accessed from: http://artsites.uottawa.ca/erplinglab (August 2016).
[2] Researchers can email ERPlinglab@gmail.com to request these Excel-based scoring templates.

monolinguals; 2) ineligible participants who should be excluded from the data analysis (e.g. due to significant exposure to a non-target language); and 3) those participants for whom more information is required. As such, only participants falling into the third category are required to complete the Extended Version of the LBQ.

For example, in a hypothetical study targeting sequential L1-English L2-French bilinguals and an English monolingual control group, the Short Version can be used to classify participants into three groups below based on the following criteria:

1) Functional Monolinguals:
   - Their native language is English, and English only; and
   - Their place of birth is an English-speaking region, and they have never lived in a non-English speaking region; and
   - The only language they have ever been immersed in is English; and
   - Their highest-attained proficiency in all other languages is low or very low.
2) Exclusions:
   - Their native language is *not* English and English only; and/or
   - Their most dominant language is *not* English; and/or
   - Their second most dominant language is *not* French; and/or
   - They have higher proficiency in a language other than in English or French.
3) More information needed:
   - Any participant not classified as Group 1 or 2.

The primary function of the Extended Version of the LBQ is to collect the detailed information needed to confirm participants' native language(s) and AoA of other languages, to operationalize MoA, language proficiency and dominance, and to assess the homogeneity of participants with respect to these quantified variables.

Defining a participant's native language(s) is not always straightforward; the first section of the Extended Version focuses on this issue by collecting data regarding exposure to languages during infancy (birth to 24 months). By having a clear picture of which languages were spoken directly to the participant by primary caregivers at which frequency, the amount and quality of naturalistic exposure to their native language(s) can objectively be assessed. The researcher can then decide if a participant should be excluded due to significant exposure to a non-target language during this crucial period of language development. For example, in the hypothetical study presented above, the researcher might want to exclude a participant who lists English as their only native language, but has a Spanish-English bilingual parent. This decision might depend on whether or not the parent actually spoke Spanish to that participant as an infant, and how much passive exposure the participant received; detailed information elicited in the Extended Version LBQ. As such, the researcher is able to make an informed decision regarding each participant's native language(s), and to assess how much heterogeneity exists across all participants, a crucial measure for grouping participants appropriately.

Operationalizing language proficiency can also be a complicated matter, as discussed in Section 3.2 of this paper. The second section of the Extended Version

focuses on collecting detailed current self-ratings of proficiency in five different skill areas (reading, writing, oral comprehension, oral production, and pronunciation) for all of a participant's languages. This provides a quick but more detailed measure of proficiency compared to the overall self-ratings provided in the Short Version of the LBQ. These ratings can be converted to a numerical value (from *very low* = 1, to *native* = 5), enabling the researcher to average ratings easily and to make comparisons across skill areas and participant groups, which also allows for correlational analyses to be made.

Perhaps the most difficult variable to operationalize is MoA. The table found in Section 4 of the Extended Version provides an objective and efficient way of doing so. The information that a participant provides here gives a detailed view of how their language use and exposure has evolved from birth to the time of testing. Participants are required to input percentages indicating their relative use of, or exposure to, each of their languages (e.g. English, French, others) in each of the defined contexts (e.g. at home, at school, etc.) and for each of the defined periods of their lives (e.g. kindergarten, elementary school, etc.). Using this information, the amount of naturalistic versus more formal exposure to the L2 can be measured in order to obtain an objective measure of MoA. To do so, each of the contexts can be assigned a weight with respect to how strongly they are considered to index "naturalistic exposure", and another weight with respect to how strongly they are considered to index "formal exposure". For example, the contexts *At home* and *Friends* might have a naturalistic weight of 1.0 and a formal weight of 0, as they are the most informal and socially active contexts, whereas the context *School* might have a naturalistic weight of 0.25 and a formal weight of 0.75 because L2 usage in that environment tends to be much less socially-based and is used primarily for instructional purposes. Weighted means for both naturalistic and formal exposure can then be calculated across each time period, and subsequently averaged by participant. Participants can then be classified into MoA groups based on their Naturalistic and Formal indices. While this might seem like a complex and time consuming procedure, a template can be used to make these calculations automatically, and can be easily tailored to reflect the particular bilingual population being tested. This table can also be tailored to provide different types of indices, for example one contrasting active and passive language exposure, which may be crucial factors in studies with heritage speakers. The data obtained from the table can also be used to assess current language dominance by looking at relative language use across different contexts during the current time frame, and it can also be used to identify when shifts in dominance have occurred throughout the lifespan.

4.2 Testing the LBQ

4.2.1 Goals

The following sections are dedicated to testing the internal validity of our LBQ and its ability to effectively operationalize variables such as AoA, Proficiency, and MoA with a diverse group of bilingual participants. A particular focus will be made on the ability of our LBQ to operationalize MoA, due to its newly emphasized importance in the L2

literature, and the fact that other language background assessment tools do not adequately characterize this variable, as discussed above.

In Section 4.2.3.1, we demonstrate the diversity and complexity of the bilingual population that is typically the source of participants for studies conducted in our lab: undergraduate students of the Ottawa-Gatineau region. Specifically, we determine whether our population is varied enough so that our groups are not confounded with respect to the variables discussed above. Variables are the most confounded if they completely overlap (e.g., if all early bilinguals are naturalistic learners with high Proficiency, then it is difficult to tease apart these three factors), and are the least confounded if a grouping with respect to one variable contains multiple sub-groups with respect to another variable (e.g., if late learners can be divided into High vs. Low Proficiency, or Naturalistic vs. Formal MoA groups).

We next verify the internal validity of our LBQ. In Section 4.2.3.2, we test whether the Short Version obtains sufficient information to reliably determine whether participants are required to complete the Extended Version. In Section 4.2.3.3, we determine whether our LBQ is able to gather complex information while not being too difficult for participants to accurately recall their language background. As such, we expect participants to be able to reproduce their own assessment of their background after an extended period of time. Specifically, if our variables (AoA, MoA, Proficiency) are well operationalized, and the questions are sufficiently precise and straightforward, the classification of participants based on those variables should be replicable.

In Sections 4.2.3.4 to 4.2.3.5, we assess the LBQ's ability to operationalize AoA, Proficiency and MoA, respectively, by determining whether they accurately confirm predictions based on what we know about these variables. For AoA, we expect AoI to be a better predictor of L2 Proficiency than Ao1E, as the latter does not entail sufficient exposure to the L2, at least in the particular case of the bilingual population tested here (see Section 3.1 for further discussion). For Proficiency, we determine whether participants are able to accurately assess themselves in different skill areas by comparing their detailed self-ratings to a more objective measure, cloze task scores. Specifically, given the strengths and weaknesses of cloze tasks discussed in Section 3.2, we expect self-assessed writing and reading skills to correlate the most with cloze test scores and the least with pronunciation. Finally, we examine whether the way we propose to operationalize MoA, as described in Section 4.1, allows us to classify participants into groups with distinct language learning situations.

4.2.2 Methods

We distributed the LBQ to participants twice, at a three month interval. The first round (test) had 81 participants, with 29 L1 French speakers, 42 L1 English and 10 non-L1 English or French speakers. The second round (retest) had 48 of the same participants, with 19 L1 French speakers, 21 L1 English speakers and 8 non-L1 English or French speakers. The non-L1 English or French speakers were excluded from all statistical analysis. During test, the participants were given both the Short and Extended version of the LBQ, as well as French and English cloze tests (Tremblay 2011; Brown 1996).

During retest, only the two versions of the LBQ were distributed. Participants were asked to record their start and end time after completing each questionnaire.

4.2.3 Results

4.2.3.1 Confounding

In this section, we verify the level of confounding for our operationalization of four variables, AoI, Ao1E, Proficiency and MoA. For each of these four variables, each participant was assigned to one of four groups for AoI, Ao1E and Proficiency, and one of 5 groups for MoA. For example, for the variable AoI, participants were classified as either *simultaneous bilinguals* (AoI = birth), *early L2 learners* (AoI < 7), *late L2 learners* (AoI > 7), or *functional monolinguals* (no AoI). As such, each participant was grouped in four different ways, assigned to a single grouping for each variable. We then analyzed each variable's groups with respect to each of the other variables and determined how much heterogeneity existed. For example, for each of the AoI groups above, we determined what proportion of participants were also classified as having High, Mid, Low or Very low L2 Proficiency. If a single AoI group was determined to consist entirely of participants from the same L2 Proficiency group, then there is no variation for that pair, meaning that the two variable groups are confounded.

Analysis of spread in each group for all our variables revealed that only four groups were completely confounded: all Simultaneous AoI participants are also Simultaneous Ao1E (SD = 0); all Monolingual Ao1E participants are also Monolingual AoI (SD = 0); all Monolingual Ao1E participants are also the lowest level of MoA (SD = 0); and all Low Proficiency participants are also the lowest level of MoA (SD = 0). This demonstrates that our particular participant pool is not very confounded, as all these pairings are very predictable; the first three pairs with Ao1E are intrinsically true, and the fact that Low Proficiency participants show very little immersion is not surprising.

4.2.3.2 Reliability

As discussed in Section 4.1, the Short Version of the LBQ gathers information regarding Ao1E, AoI and overall proficiency, and we discuss here whether this data is sufficient to reliably determine whether or not participants need to also complete the Extended Version. To do so, we grouped participants in three ways: 1) *functional monolinguals*; 2) *exclusions*, participants who are ineligible for a typical bilingualism study due to a non-target L1, and 3) *bilinguals*, participants to whom we would give the Extended Version under normal circumstances. We then used the additional data provided in each participants' Extended Version to classify them again, and compared the classifications from the two versions. We found that participants were grouped consistently 100% of the time, which indicates that the data provided in the Short Version is sufficiently detailed in order to accomplish this primary objective.

4.2.3.3 Replicability

In order to determine whether the responses to the questions in the LBQ could be replicated by participants over time, we gave the LBQ to participants twice at a three month interval (test and retest). We then compared values for each grouping variable for each participant that was present at both sessions. These results are presented in Figure 1, where each data point represents a participant's pair of grouping values for a given variable at test and retest.

A paired-samples t-test on grouping values revealed no significant difference between test and retest, when all four variables were collapsed ($t(184) = 0.419$, $p = .676$). This demonstrates that the questions targeting these variables are sufficiently straightforward for participants to provide accurate and reliable responses.

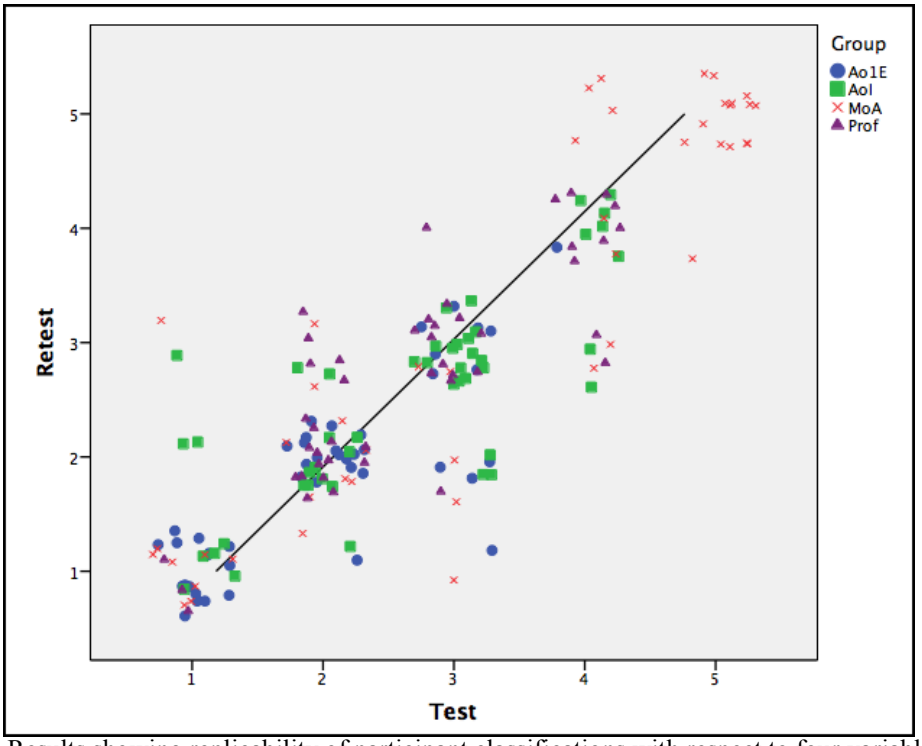

Figure 1. Results showing replicability of participant classifications with respect to four variables at test and retest. *Note: points are jittered, significant correlation n=185, r=0.882, p<.001*

Analyses on individual variables revealed no significant differences at test and retest for AoI, Proficiency, or MoA (all p's > .323). However, there was a significant difference for Ao1E ($t(45) = 2.209$, $p < .05$), such that participants tended to be classified in a higher age group at the test phase (mean = 1.94, SD = .83) than at the retest phase (mean = 1.8, SD = .78). We investigated this effect further by comparing absolute age values provided by participants for Ao1E (e.g. 4 yrs old) at test and retest, rather than comparing the Ao1E group they were placed in (e.g. *Early L2 learner*). This paired-samples t-test revealed no significant difference between the test (mean = 4.67, SD = 3.44) and retest (mean = 4.58, SD = 3.73) for Ao1E ($t(47) = .319$, $p = .751$). This suggests that participants are actually quite consistent at giving their age values, but that any existing

discrepancies likely occur around group classification boundaries. This highlights the flaws of using continuous variables, such as age, as categorical variables, as is often done in L2 studies.

4.2.3.3 Operationalizing AoA

In order to verify the importance of distinguishing between AoI and Ao1E when operationalizing AoA, we ran a multiple regression analysis to predict self-rated proficiency based on AoI and Ao1E. The predictors explained 28% of the variance (R^2 = .28, F(2,62) = 12.026, p < .001). The analysis shows that, as expected, AoI did significantly predict the value of Proficiency, (R^2 = .001, t(64) = -3.67, p < .05); however, Ao1E did not significantly predict the value of Proficiency (R^2 = .276, t(64) = .029, p = .977). This demonstrates the value of choosing to operationalize AoA as the age of L2 immersion in populations such as the one sampled here.

4.2.3.4 Operationalizing L2 proficiency

This section has two goals. The first is to demonstrate that the Short Version of the LBQ provides an adequate coarse-grained measure of L2 proficiency. To do so, we ran a correlation analysis between participants' self-rating for "Current level of general proficiency" and their corresponding cloze task score, which we assume to be an objective measure of overall proficiency. This revealed a significant correlation (r(83) = 0.667, p < .001), confirming our hypothesis that participants are generally accurate when self-assessing their L2 proficiency. We also ran an additional correlation analysis between this same self-rating score from the Short Version, and the average of the detailed self-rated scores provided in the Extended Version. Unsurprisingly, results indicate that this correlation is also significant (r(85) = .336, p =.002), further demonstrating that this "general level" self-assessment is a reasonable proxy for their overall level of L2 proficiency, and also that participants' responses are consistent between Short and Extended Versions.

The second goal is to demonstrate that there is a benefit to collecting detailed self-ratings in different skill areas. To do so, we ran multiple correlations to determine which language skill area(s) best correlated with different variables. Results inidicated that L2 Cloze task scores correlate most with Self-rated reading proficiency (r(83) = .722, p < .001), and least with Self-rated pronunciation (r(83) = .496, p <.001). This corresponds to our expectations based on the suitability of cloze tasks to evaluate these aspects of proficiency. The results also demonstrated that Ao1E and AoI correlate most with Self-rated pronunciation (r(79) = -.395, p < .001, r(64) = -.536, p < .001, respectively), which is another intuitive finding, as critical/sensitive period effects are most robustly attested for phonological aspects of L2 acquisition. Finally, it was also determined that MoA correlates most with overall Self-rated proficiency (r(75) = .749, p <.001), and the least with Self-rated writing (r(76) = .545, p <.001) and reading (r(76) = .563, p <.001), signifying that a parallel can be made between naturalistic learning and simultaneous learning, where communicative skills are being developed more than formal ones.

Overall, the results here indicate that participants are able to provide detailed self-ratings that reflect the complexities of assessing language proficiency.

4.2.3.5 Operationalizing MoA

In order to determine whether or not the data collected in the table found in Section 4 of the Extended Version is able to appropriately categorize participants into relevant groups, a cluster analysis was performed. Cluster analysis is an exploratory statistical method that is used to classify a sample of participants into sub-groups based on their similarities and differences with respect to a set of measured variables. In this case, the measured variables that were used were the Naturalistic and the Formal Indices that were calculated for each of the 49 participants who had English as an L1 (including seven simultaneous bilinguals who were classified as L1-French in Section 4.2 due to dominance in French).

Specifically, the Naturalistic Index was calculated by assigning the following factor weights to the nine different contexts listed in the table: 1.0 for *Family*, *Friends*, *Work* and *Partner*; 0.5 for *Daily activities*, *Extended family*, *Media* and *Reading*; and 0.25 for *School*. For each time frame, these factor weights were then multiplied by the value that the participant had entered in that cell for their use/exposure to French (the L2), and a weighted mean was calculated. These weighted means were then averaged across all time frames to produce the Naturalistic Index, expressed as a percentage. To calculate the Formal Index, the context *School* was assigned a factor weight of 1.0, and all other factors were assigned a factor weight of 0, reflecting the fact that there is really only one environment where formal exposure to French occurs (i.e., in the classroom). As such, the value that the participant had entered for their use/exposure to French in the *School* context was averaged across all time frames to produce the Formal Index, also expressed as a percentage.

Participants' values for these two variables were entered into a hierarchical cluster analysis using between-group linkages, with participants joined into clusters using squared Euclidian distance (a measure of how different a participant's Naturalistic and Formal Index is). Results revealed that the participants could be classified into four distinct groups based on these two variables, shown below in Table 1.

Table 1. Participant groups based on hierarchical cluster analysis.

#	***Group***	***n***	***Naturalistic Index (%)***		***Formal Index (%)***	
1	"Monolinguals"	21	**Mean: 2.5** (2.3) Range: 0.0-7.8	LOW	**Mean: 8.4** (7.8) Range: 0.0-17.0	LOW
2	Fr-Immersion	18	**Mean: 8.5** (6.2) Range: 1.6-26.8	LOW-MID	**Mean: 30.9** (7.8) Range: 21.0-47.0	MID
3	EN-Dominant bilinguals	3	**Mean: 36.7** (1.3) Range: 35.1-38.7	MID	**Mean: 16.3** (2.1) Range: 14.0-18.0	LOW
4	FR-Dominant bilinguals	7	**Mean: 62.2** (19.0) Range: 32.4-83.5	HIGH	**Mean: 81.5** (9.6) Range: 65.0-93.0)	HIGH

The first group consisted of 21 participants who had low values for both the Formal Index (mean = 8.4%, SD = 7.8%) and the Naturalistic Index (mean = 2.5%, SD = 2.3%). As such, these appear to be typical English (functional) monolinguals who have had minimal exposure to French in a formal setting, and even less naturalistic exposure. The second group consisted of 18 participants who had a moderate Formal Index (mean = 30.9%, SD = 7.8%) and a relatively low Naturalistic Index (mean = 8.5%, SD = 6.2%), which is typical of those who learned French as an L2 in an immersion program.

The last two groups that emerged consisted almost entirely of simultaneous bilinguals (the one exception had an AoI/Ao1E of 4). The first of these groups consisted of three participants who had a relatively low Formal Index (Mean = 16.3%, SD = 2.1%) and a moderate Naturalistic Index (mean = 36.7%, SD = 1.3%), which seems to be consistent with English-dominant bilinguals who have perhaps undergone some attrition of French as a result of being in an English school system in an English community and thus forging social relationships based in English. Note that it would be uncommon for native speakers of French to be enrolled in French Immersion programs, which are much more geared towards L2 learning, not L1 maintenance. The second of these two groups of simultaneous bilinguals (and one early L2 learner) consisted of seven participants with a high Formal Index (mean = 81.5%, SD = 9.6%) as well as a high Naturalistic Index (mean = 62.2%, SD = 19.0%). These are more typical French-dominant bilinguals who have attended French-language schools in French-language communities, and as such have formed social ties based in French.

The results here indicate that the hierarchical cluster analysis was indeed successful at separating the participants into distinct groups with meaningful characteristics based on their similarities and differences with respect their MoA. Because MoA was operationalized using the values for the Naturalistic and Formal Indices that were calculated directly from the data provided in the LBQ, this suggests that the LBQ provides a valid means of evaluating this complex variable.

5. Conclusions

In this paper we have discussed the main variables (specifically AoA, Proficiency, Dominance, and MoA) that contribute to the potentially complex language backgrounds of the participants in many bilingualism and second language acquisition studies. In order to tease apart these variables, we propose that our comprehensive LBQ is capable of assessing and quantifying the multiple factors that affect bi- and multilingual processing. This tool improves upon other currently available language questionnaires in its completeness and its ability to systematically tease apart the variables.

Our LBQ includes a Short Version which can be administered quickly in the hopes of confirming that participants are either functional monolinguals (and thus not needing to fill out the more complex Extended Version), are participants with potentially more complicated language backgrounds (thus needing to fill out the Extended Version), or do not qualify for the particular study being run. This capability is beneficial in that it reduces the time for filling out the LBQs and it has been shown to be 100% reliable in

making this first pass gross participant grouping. We have further shown that the participant groupings formed in the interpretation of the LBQ can be replicated after a three month lag. Finally, we have also demonstrated that we can accurately confirm predictions based on what we know about AoA, Proficiency and MoA, and thus the LBQ is successful at allowing researchers to operationalize these variables.

Based on the results presented here and with the goal of making improvements to how we have operationalized MoA (a newer variable in the field of the Psycholinguistics of bilingualism), we are currently improving the MoA table in our LBQ. These additions will enable researchers to make finer-grained MoA groups that may better fit any given participant population and are more aligned with the predictions in the field of L2 and bilingual processing.

References

Brown, James Dean. 1996. *Testing in language programs*. Upper Saddle River, NJ: Prentice Hall.

Birdsong, David. 1992. Ultimate attainment in second language acquisition. *Language* 68: 706-755.

Dussias, Paola E., and Nuria Sagarra. 2007. The effect of exposure on syntactic parsing in Spanish–English bilinguals. *Bilingualism: Language and Cognition* 10(1): 101–116.

Johnson, Jacqueline S., and Elissa L. Newport. 1989. Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology* 21: 60–99.

Lenneberg, Eric H. 1967. *Biological Foundations of Language*. New York, NY: Wiley.

Li, Ping, Fan Zhang, Erlfang Tsai, and Brendan Puls. 2014. Language history questionnaire (LHQ 2.0): A new dynamic web-based research tool. *Bilingualism: Language and Cognition* 17(3): 673-680.

Marian, Viorica, Henrike K. Blumenfeld, and Margarita Kaushanskaya. 2007. The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech Language and Hearing Research* 50(4): 940-967.

Pliatsikas, Christos, and Theodoros Marinis. 2013. Processing empty categories in a second language. When naturalistic exposure fills the (intermediate) gap. *Bilingualism: Language and Cognition* 16: 167-182.

Sabourin, Laura, Christie Brien, and Michèle Burkholder. 2014. The effect of age of L2 acquisition on the organization of the bilingual lexicon: Evidence from masked priming. *Bilingualism: Language and Cognition* 17(3): 542-555.

Sabourin, Laura, Jean-Christophe Leclerc, Michèle Burkholder, and Christie Brien. 2014. Bilingual lexical organization. Is there a sensitive period? *Proceedings of the Canadian Linguistic Association*, ed. Laura Teddiman. http://cla-acl.ca/wp-content/uploads/Sabourin_Leclerc_Burkholder_Brien-2014.pdf

Treffers-Daller, Jeanine. 2016. Language dominance: The construct, its measurement and operationalization. In *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*, ed. C. Silva-Corvalan and Jeanine Treffers-Daller, 235-265. Cambridge: Cambridge University Press.

Tremblay, Annie. 2011. Proficiency assessment standards in second language acquisition research: "Clozing" the gap. *Studies in Second Language Acquisition* 33: 339-372.