**Appendices**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No.** | **Name of Journal and Its Writer** |  | **Modality Value** |  |
|  |  | **High** | **Middle** | **Low** |
|  |  |  |  |  |
| 1. | Journal of Memory and Language (JML) (Robey, 2019) |  | Concerns regarding these differences **should** be minimal, as Abel and Roediger (2017) have shown that the testing effect is equivalent across blocked and intermixed study designs.  All results regarding this measure, however, **should** be taken cautiously.  More work is needed to determine how to best conceptualize how variability in strategy use **should** be measured before it becomes widely used within this domain.  Future research hoping to explore strategy variability using the CWS ratio **should** explore how much exposure to a particular task is needed for strategy expertise to develop.  Special note **should** be made, however, for how this measure was constructed, as similar measures of shallow and deep strategy use from Minear et al. (2018) were not significant.  It **should** be noted, however, that Minear et al. (2018) did find significant results related to shallow and deep strategy use, but only for certain items or when the testing effect was treated as a categorical rather than continuous variable. | Results regarding whether this new measure **could** be used to predict the testing effect were largely inconclusive and SEM analyses were all non-significant. |
| 2. | Language, Culture and Curriculum (LCC) (Bower, 2019) |  | In a national context of prevailing demotivation for language learning, senior leaders, it **would** appear,  Senior leaders it **would** seem are indeed instrumental to sustaining CLIL in an innovation-limiting culture of accountability.  The next section **will** discuss affordances and constraints arising from issues that emerged from the data.  This **would** seem to contradict concerns raised in other contexts (e.g. Bruton, 2011, 2015; Van Mensel et al., 2019)  It **will** be interesting to see whether within the macro alliances of the new self-improving school system, leaders from such innovations do cross boundaries so that cross-curricular language learning is enabled to develop across England as one curriculum initiative.  To be successful, these programmes **would** need to be supported by colleagues with a cogent grasp of the pedagogical theory: understanding why something works is critical if it is to be further developed effectively both in-house and across institutional boundaries.  It **would** be particularly interesting to further investigate CLIL contexts in which motivation to learn a language is traditionally problematic, e.g. Anglophone-dominant contexts, particularly those with compulsory models. | …view their CLIL programme as part of the solution to the raising attainment agenda whilst acknowledging some of the challenges that introducing CLIL **can** create.  It **may** originate in strategic goals (exam success) as well as societal goals but it appears to be sustained by the perceived pedagogical benefits to learners generated by learning in this way.  At the meso level, leadership was well established in each context. It is interesting that head teachers and a senior leader in each context understood the impact on school improvement that CLIL was making, or **could** make (Cedar School), and fully supported the programme.  Without this support it **may** not be possible to implement and to sustain innovations in the face of pressures such accountability measures, the narrow curriculum, initial resistance from staff, pupils or even parents.  that selection **may** play a significant role in higher outcomes. In the short term, to breach the gap in pedagogical knowledge, teacher educators may have an important role to play in introducing CLIL methodology and practice.  Institutions within the developing self-improving system **may** potentially draw on such resources.  Interestingly, interviewees were confident of CLIL’s positive impact on learner progress, without apparently fully understanding how and why this **might** be achieved.  Perhaps this lack of pedagogical understanding in leaders **may** have contributed to the slow uptake of CLIL in England compared to that elsewhere in Europe and to its demise when key staff move on.  In a school development system in which training comes from within, such lack of pedagogical understanding **may** be of greater significance.  Challenge to professional integrity (Moate, 2011) that **can** demotivate teacher practice was not evident in the findings.  This **may** be because the few teachers involved were confident in this approach and had opted to be involved.  This **may** relate to the nation-specific CLIL profile of England– in these schools, learners of all abilities enter their language national exams at least one year earlier than their peers and revert to English for the curriculum subject. |
| 3. | Brain and Language (BL) (Bice & Kroll, 2019) |  | However, if SES were the primary factor, then Pennsylvania monolinguals **should** have outperformed the California monolinguals.  … other consequences for our participants (e.g., access to resources or opportunities) that **would** only have been biased in favor of the Pennsylvania students.  Many studies **would** exclude the monolinguals from California in our sample ….  **Would** the California monolinguals be better able to learn a new language in a classroom setting?  In future research it **will** be of great interest to determine whether openness to new learning is created by immersion in a diverse environment itself or whether specific features of the context or language experience are critical.  In retrospect, our conclusions **would** be stronger if we had been able to limit the California sample of monolinguals to only those who had been living in the context for most or all of their lives;  Nevertheless, if our sample did include monolinguals who had only recently moved to that context, then the effect **would** likely be smaller, leading to greater noise in the signal or an underestimation of the effect.  Likewise, we **would** have liked to have been able to ask more detailed questions regarding language experiences, exposure to languages other than those that the participants reported having previously studied, performance in foreign language classes, socioeconomic status, etc. | The observation that the ERP deflection was found in the real words rather than to the violations **may** imply that the California monolinguals learned something about what forms the real words, rather than something about what made the “fake” words different.  There are potential factors that it was not possible to control in the current study that **could** have impacted the results, including SES.  Therefore, in many ways, one **might** expect larger differences between the monolingual groups based on SES factors alone, in favor of those living in Pennsylvania.  Although SES differences have not been directly linked to second language learning or learning more generally, particularly in college-age students, the differences in SES **may** have had other consequences for our participants ….  The ERP differences between the two monolingual groups suggests that ambient linguistic diversity, and the various other factors co-involved in linguistic diversity, **may** have a positive impact on new language learning.  As noted, there are other factors that **may** have contributed to these differences.  Even if ambient linguistic diversity is the factor driving this effect, we **cannot** tell from these data alone whether it is simply overhearing foreign languages, the requirement to interact regularly with non-native speakers of English, or actual differences in past language learning experience.  Living among linguistic diversity not only increases the number of interactions with accented speakers, but also **may** have consequences for the regional dialect.  … for not being what one **might** consider “pure” monolinguals.  The effects of ambient linguistic diversity **may** prepare individuals for some, but not all, aspects of new language learning.  However, most studies on language immersion examine late bilinguals who are intentionally learning the language and who are moving from a relatively unilingual context into an immersion context that **may** itself be unilingual or linguistically diverse.  This lessened our experimental control and ability to report many variables and factors that **could** have been useful for disentangling the source of the effects.  However, despite the lack of empirical measures to describe how these groups **may** differ, we fully believe that the differences between contexts and the effects on language learning reported here are true and representative. |
| 4. | Language, Cognition and Neuroscience (LCN) (Kaiser, 2019) | These results suggest that language users separately track expectations about who **is likely to be** mentioned next and what kind of event **is likely to be** mentioned next. |  | Prior linguistic work leaves open the question of whether the prominence of entities and events in mental models of discourse **can** be dissociated. |
| 5. | Journal of Child Language (JCL) (Jones & Brant, 2019) | A limitation of the current study is that **it is impossible to** provide evidence for any causal account on the basis of correlational data alone.  What the current study shows is that any explanatory model of early vocabulary development, particularly of early word production, **must** account for word sound features. |  | A word with such characteristics but complex phonology **may** be memorised imperfectly, which may be sufficient if the child is required to recognise and respond to, though not necessary produce, such a word (e.g., “Eat your breakfast!”  As a result, young children **may** tend implicitly towards acquiring new words comprising familiar phonological patterns. |