



Does verbatim sentence recall underestimate the language competence of near-native speakers?

Judith Schweppe^{1*}, Sandra Barth², Almut Ketzer-Nöltge¹ and Ralf Rummer¹

¹ Department of Psychology, University of Erfurt, Erfurt, Germany

² Project for Successful Teaching and Studies, Kiel University, Kiel, Germany

Edited by:

Matthew W. Crocker, Saarland University, Germany

Reviewed by:

Judith Koehne, Bamberg University, Germany

Heiner Drenhaus, Saarland University, Germany

*Correspondence:

Judith Schweppe, Department of Psychology, University of Erfurt, Nordhäuser Str. 63, 99089 Erfurt, PO Box 900 221, 99105 Erfurt, Germany
e-mail: judith.schweppe@uni-erfurt.de

Verbatim sentence recall is widely used to test the language competence of native and non-native speakers since it involves comprehension and production of connected speech. However, we assume that, to maintain surface information, sentence recall relies particularly on attentional resources, which differentially affects native and non-native speakers. Since even in near-natives language processing is less automatized than in native speakers, processing a sentence in a foreign language plus retaining its surface may result in a cognitive overload. We contrasted sentence recall performance of German native speakers with that of highly proficient non-natives. Non-natives recalled the sentences significantly poorer than the natives, but performed equally well on a cloze test. This implies that sentence recall underestimates the language competence of good non-native speakers in mixed groups with native speakers. The findings also suggest that theories of sentence recall need to consider both its linguistic and its attentional aspects.

Keywords: sentence recall, bilingualism, near-native speakers, attention, language competence, working memory

INTRODUCTION

Verbatim sentence recall (or sentence repetition testing, Diller and Jordan-Diller, 2003) is a task widely used in tests of language proficiency in educational (e.g., Grimm, 2001; Fried, 2008) and clinical contexts (e.g., Meyers et al., 2000) because it discriminates well between good and less good performers (Grimm, 2001). It has been used for measuring second language (L2) competence (Radloff and Hallberg, 1991; Diller and Jordan-Diller, 2003) and is also included in language development tests (e.g., for German: SETK 3-5, Grimm, 2001; SSV, Grimm, 2003; for English: TOLD-P:4, Newcomer and Hammill, 2008; KLST-2, Gauthier and Madison, 1998). An advantage of using sentence recall in language proficiency tests is that it can be conducted with little effort. Furthermore, verbatim sentence recall covers many aspects of language processing: it requires comprehension and production skills, and involves processing at phonological, lexical-semantic, morphosyntactic, syntactic, and propositional levels (e.g., Schweppe, 2006). It is therefore also used in psycholinguistic research for studying syntactic priming (e.g., Potter and Lombardi, 1998; Meijer and Fox Tree, 2003).

In spite of these advantages, we suggest that the verbatim recall of sentences is not that good a measure for estimating differences in language competence between native and highly proficient non-native speakers, for sentence recall systematically underestimates language proficiency of these highly proficient L2 speakers. In online sentence processing, there is usually no need to maintain surface information. However, we assume that when a sentence is processed for verbatim recall, surface representations need to be kept available, a process that is cognitively costly (Aaronson and Scarborough, 1976; Rummer et al., 2013). The need to recall the exact wording of sentences would thus not only require verbal

competence but also a substantial amount of general attentional resources. When, in addition to the maintenance, also the processing of the sentence itself is cognitively costly—as is the case in a non-native language, even when language competence is very high (Clahsen and Felser, 2006)—the attentional demands imposed by verbatim sentence recall could be too high and recall performance could thus break down. Consequently, if language competence of (highly) proficient non-native speakers is evaluated based on verbatim sentence recall, it could be underestimated as compared to other tasks and as compared to native speakers whose language processing is more automatized. This should even be the case for non-native speakers with native-like proficiency in other linguistic tasks—so-called *near-natives*—whose performance on sentence recall would then nevertheless be considerably poorer than that of natives.

In light of these considerations, a specific application of sentence recall appears critical: The task is a core part of an annual language screening for preschool children in Germany in which children with special needs are identified (e.g., “DELFIN 4,” Fried and Briedigkeit, 2007; Fried, 2008). Crucially, the assessment of the test outcomes does not consider whether the children are L1 or L2 speakers. The most prominently discussed outcome of these language screenings was the finding that children with German as an L2 performed dramatically poorer than L1 children. If our assumptions are correct, this finding could partly be due to sentence recall underestimating the language proficiency of good L2 speakers as compared to native speakers. Consequently, one should avoid lumping together native and non-native speakers when using this task.

In the present paper, we test our assumptions by investigating (adult) native and highly proficient non-native speakers of

German with respect to their performance in a verbatim sentence recall task. In addition, we test their performance in a complex language task that also addresses comprehension and production skills but does not require explicit maintenance and which should thus not demand substantial attentional resources. However, from the perspective of the *conceptual regeneration hypothesis* (Potter and Lombardi, 1990), one could question the assumption that verbatim sentence recall requires the active maintenance of surface information and that processing a sentence for verbatim recall is therefore attentionally more demanding than processing it for comprehension. We will thus briefly describe Potter and Lombardi's (1990, see also Lombardi and Potter, 1992; Potter and Lombardi, 1998) account as well as our modification (Rummer et al., 2013).

According to the conceptual regeneration hypothesis, sentences are not stored in working memory but are regenerated based on the propositional representations abstracted during comprehension. Sentence recall is therefore simply a combination of sentence comprehension and sentence production. Nonetheless, recall is often verbatim, since lexical-semantic representations of the words in the sentence are activated in the course of comprehension, and participants access these during regeneration with a higher probability. Rummer and colleagues (e.g., Rummer and Engelkamp, 2001, 2003; Rummer et al., 2003; Schweppe and Rummer, 2007; Schweppe et al., 2011) modified Potter and Lombardi's (1990) approach such that presentation of a (to-be-comprehended or to-be-recalled) sentence automatically activates long-term memory representations on all levels of language processing and that these representations can be kept available. Which representations are maintained depends on the affordances of the particular task. Surface representations, such as phonological ones, are kept available only when the task is to recall the exact wording of a sentence (Rummer et al., 2013). The mechanism responsible for prolonging activation of otherwise dispensable surface representations is related to the allocation of attention and is assumed to be cognitively demanding¹. This idea is based on more general theories that think of working memory as attentional processes operating on long-term memory (e.g., Cowan, 1999; Barrouillet et al., 2004). Cowan's *embedded processes model* (Cowan, 1995, 1999, 2001) assumes that during encoding a stimulus activates multiple features in long-term memory and that their activation can be maintained by directing attention inwards to representations that are relevant for the current task. Attention is a limited resource that is shared between processing and maintenance of activation. Verbatim sentence recall should thus be attentionally more demanding than

sentence comprehension or gist recall, as more representations need to be maintained (Rummer et al., 2013). A reading time study by Aaronson and Scarborough (1976) suggests that this is indeed the case: when reading a sentence for recall, reading times were significantly higher than when participants read the same sentences for comprehension. Another finding that is in line with the idea that sentence recall demands attention comes from a study by Baddeley et al. (2009). They investigated (auditory) sentence recall as a single task and in combination with a visual continuous reaction time task that required general attention and observed poorer recall performance in the dual task condition than in the single task condition.

The idea that attention is shared between processing and maintenance also suggests that the more attention demanding a processing task is, the more it should hinder maintenance (e.g., Cowan, 1999; Barrouillet et al., 2004). One factor that affects the attentional demands of language processing is whether a native or a non-native language is processed: even in non-natives who command their L2 very much like natives and perform native-like in other language tasks, language processing is less automatized than in an L1 (e.g., Bosch et al., 2000; Clahsen and Felser, 2006). For instance, Stowe and Sabourin (2005) report that although the same brain areas are active during sentence comprehension (and other processing tasks) in L1 and L2, the activation when processing an L2 is increased even in highly proficient speakers with age of acquisition around the age of three. Similar findings are reported for Spanish-Catalan bilinguals who were exposed to both languages to the same degree from their third year on (Perani et al., 2003). Importantly, the increase of activation during L2 processing was observed even in the absence of performance differences between L1 and L2 speakers (see also Birdsong, 2006). A meta-analysis of neural processing differences between native and non-native speakers suggests that these can most reliably be found in the left prefrontal cortex, specifically in the left inferior frontal gyrus (e.g., BA 47), which is involved in non-lexical compositional processes such as syntactic processing during sentence comprehension (Indefrey, 2006). The same areas are also involved in executive processes such as attentional control (Miller and Cohen, 2001). Higher activation levels of prefrontal areas in non-native sentence processing may thus either reflect compensation for lower efficiency in these regions (Indefrey, 2006) or "executive control over access to short- or long-term memory representations" (Abutalebi, 2008, p. 472; see also Thompson-Schill et al., 1997; Fletcher et al., 1998) and thus higher attentional load.

If (1) non-native language processing is indeed attentionally more demanding than native language processing even in highly proficient bilinguals, and (2) verbatim sentence recall is a particularly demanding verbal task, the need to recall the exact wording of a sentence in a non-native language may overload the attentional system. This kind of overload is less likely to occur in L1 sentence recall because in this case attentionally demanding maintenance is combined with attentionally less demanding processing. The imperfect automaticity of highly proficient L2 speakers, which does not have consequences for performance in most linguistic tasks, would lead to substantial consequences in an attentionally demanding task such as sentence recall. In other words, verbatim sentence recall should increase performance

¹There is a discussion going on whether forgetting in working memory is due to a decay of activation that can be prevented via rehearsal (e.g., Barrouillet and Camos, 2009) or whether it is due to interference between competing representations (e.g., Lewandowsky et al., 2009). We talk about the maintenance of activation in this context and refer to a working memory model that emphasizes the role of decay. However, the basic assumptions apply irrespective of whether one assumes that it is attentionally demanding to prevent activation from decay and thus to keep the activated representations available for verbatim recall or that it is attentionally demanding to reduce interference from competing representations and thus to prevent intrusions from lexical competitors or syntactic alternatives during verbatim sentence recall.

differences between native and non-native speakers as compared to a task that also taps comprehension and production skills at sentence or discourse level but that does not pose additional maintenance demands.

The present study tests this assumption by comparing native and non-native speakers' performance in sentence recall and on the C-Test, which follows the principles of a cloze-test and consists of short texts with gaps (Raatz and Klein-Braley, 1982). Like sentence repetition, C-test performance relies on both sentence comprehension and sentence production and taps semantic as well as syntactic skills. However, unlike sentence repetition, it does not require the controlled maintenance of surface representations. We therefore predict that when comparing a group of highly proficient non-natives and a group of natives, performance differences will be considerably larger in sentence recall than on the C-test. A particularly strong test of this hypothesis would be to compare natives and non-natives with similar C-test performance. We thus contrasted non-native participants with high scores on the C-Test and a sample of native speakers of German. In addition, we tested participants' performance in sentence recall with both auditory and visual presentation since it was an open question whether one of the two input modalities might cause particular difficulties for the near-native speakers.

STUDY

The experiment was based on a 2×2 design with the quasi-experimental variable language group (native speakers of German vs. near-native speakers of German) and the within-subjects variable presentation modality (auditory vs. visual). Performance on the C-Test and in verbatim sentence recall served as dependent variables. Furthermore, the C-Test score was included as a covariate in the analyses on sentence recall, and sentence recall performance was included as a covariate in the analyses for the C-Test.

METHOD

Participants

Seventy eight participants, 54 native, and 24 non-native speakers of German, were tested. The non-native speakers (mean age: 26.67 years, range 18–60) were students or employees at the (German) universities of Erfurt, Hamburg, Leipzig, and Saarbrücken. They all reported to have had at least 5 years of experience with German as a second language (mean number of years of experience = 14.2 years) and had participated successfully in German school or university education or both. Each non-native participant reported to have the German general qualification for university entrance. Their long stay in Germany and their successful educational history indicate that the subjects mastered their L2 to an extremely high degree. The non-native group was heterogeneous with respect to their first language (e.g., Indonesian, Finnish, Serbo-Croatian). A broader sample of native participants (with respect to age and educational background) was tested². The pool of the native speakers (mean age: 19.09 years, range 13–27) included 20 students at the University of Erfurt (18–25 years of age), 15 high

school students attending a school of the college-bound track of the German school system ("Gymnasium") in Erfurt (at the transition from grade 9–10; 13–14 years of age) as well as 19 vocational school students (18–27 years of age), 12 of whom were also studying for a qualification for university entrance.

Materials and procedure

Participants had to conduct two tests: first, a paper-pencil based C-Test, and second, a computer-based verbatim sentence recall task.

C-Test. The C-Test was first developed by Raatz and Klein-Braley (1982) and has since been studied in many languages (for an overview see the C-Test bibliography by Grotjahn, 2007). Native speakers' C-Test scores correlate with school grades in native language classes up to grade eight (Wockenfuß and Raatz, 2006). For L2 learners, C-Test scores correlate highly with performance in second language courses (Eckes, 2010) and are widely used to assign learners to course levels. Furthermore, C-Test scores correlate highly with scores on institutionalized L2 proficiency test batteries, for instance with the TOEFL ($r = 0.55\text{--}0.91$), TOEIC ($r = 0.62$), and the Michigan Test ($r = 0.54\text{--}0.61$), and with the Oxford Placement Test ($r = 0.83$) in English and the TestDaF ($r = 0.76$) in German (for an overview see Eckes and Grotjahn, 2006). For the purpose of our study it may be problematic that older adolescent and adult native speakers should score at ceiling on most C-Tests. However, it is possible to construct C-Tests so that even the native speakers do not score perfectly (Baur and Meder, 1994). The C-Test used in our study consisted of four short texts with 99 gaps (for the original texts and an English translation see the Appendix in Supplementary Material). The texts differed in their theme and style, their difficulty increased. In line with the standard C-Test procedure, 5 min were given for each short text. In general, participants took less than 5 min to complete each of the texts. We chose the texts to be sufficiently difficult to induce errors by the native speakers. Only one of the 54 natives (and one of the 24 non-natives) achieved a perfect score, which demonstrates that this was indeed the case in our sample.

Sentence recall. After completing the C-Test, subjects received instructions for the computer-based part of the study. Instructions were presented on a computer screen and subjects were encouraged to ask questions when the procedure was not clear to them. The sentence recall task consisted of 40 sentences, which were adapted from the German version of Daneman and Carpenter's (1980) reading span test (Hacker et al., 2002). Since native speakers are able to recall sentences of up to 16 words or more (Brener, 1940), the sentences used here were slightly modified in length to induce errors in L1 sentence recall. Each modified sentence included 16–20 words (mean length 17.8; $SD = 1.12$). The materials (including the audio files) are available for download at the IRIS digital repository (<http://www.iris-database.org/iris/app/home/detail?id=york%3a815588> and). Sentences were divided into four blocks with mean sentence length of 17.9, 17.7, 18, and 17.5 using a latin square technique. The sentences were recorded for auditory presentation by a female native speaker. In the visual condition, they were presented center screen white on black with font Arial and size

²This was originally done to increase the probability of identifying native speakers who could be matched with respect to their sentence recall performance.

22 pt. Sound files lasted between five and eight seconds and presentation times of the written sentences matched presentation times of the auditory versions. The four blocks of sentences were balanced across conditions, each participant was presented two blocks, and no participant was presented the same sentence more than once. A *post-hoc* analysis revealed no differences in the ease of recall between the four blocks. In the first block, sentences were presented auditorily via headphones. The second block was presented visually. This order was the same for all participants. Within blocks, the order of the sentences was randomized. After each sentence was presented, the participants had time as needed to write down the exact wording, and then they started the next trial by pressing the space bar. Each block began with two practice trials. Finally, participants completed a questionnaire regarding their language history, were debriefed and received payment or course credit for their participation.

RESULTS

First, we will present analyses with sentence recall performance as the dependent variable, which will be followed by the analyses with C-Test performance as dependent variable. One native speaker (C-Test score 99; (auditory) recall performance 97.28%) and one non-native speaker (C-Test score 97; (auditory) recall performance 82.32%) only completed the auditory items. To base the analyses for both dependent variables on the same pool of participants, we also excluded these two participants from the analyses on C-Test scores.

Sentence recall

We scored participants' responses according to a strict criterion but disregarding word order. Words were scored as correct only when they occurred in the same grammatical form as in the original sentence, while purely orthographic mistakes were ignored. We subjected the proportion of correctly recalled words per sentence to an ANCOVA with "language" as between-subjects variable, "modality" as within-subject variable as well as age and C-Test score as covariates.

As expected, there was a significant main effect for language in that the L1 speakers (79.52% correct, $SE = 1.52$) outperformed the L2 speakers [69.9%, $SE = 2.48$; $F_{(1, 72)} = 9.61$, $p = 0.003$, $\eta_p^2 = 0.12$]. There was also a significant effect for the covariate C-Test score [$F_{(1, 72)} = 34.41$, $p < 0.001$, $\eta_p^2 = 0.32$], while age did not significantly influence recall performance ($F < 1$).

We varied the input modality in the sentence recall task in order to explore whether one of the two modalities caused particular difficulties for the near-native speakers. As there was no interaction between language and modality ($F < 1$) and both groups performed better with auditory than with visual presentation (77.29% correct, $SE = 1.39$, vs. 72.12% correct, $SE = 1.52$; $F_{(1, 72)} = 5.65$, $p = 0.02$, $\eta_p^2 = 0.07$), this does not seem to be the case.

C-Test

For the C-Test, we awarded one point for each correctly filled gap such that the maximum score was 99. We subjected the resulting C-Test scores to an ANCOVA with "language" as between-subjects variable, and age and sentence recall performance (proportion of correctly reproduced words) as covariates.

There was a significant effect for the covariate sentence recall performance [$F_{(1, 72)} = 34.36$; $p < 0.001$, $\eta_p^2 = 0.32$], whereas the covariate age only approached significance [$F_{(1, 72)} = 3.34$; $p = 0.07$, $\eta_p^2 = 0.04$]. Crucially, the main effect for language did not reach significance ($F < 1$). L2 speakers scored as high on the C-Test (91.09, $SE = 0.93$) as did the L1 speakers (90.43, $SE = 0.56$).

DISCUSSION

The experiment aimed at demonstrating that sentence recall underestimates the language proficiency of very good L2 speakers as compared to a similarly complex verbal task that does not pose additional attentional demands. As predicted, near-natives with a highly successful educational history, long stay in Germany, and hardly any foreign accent in German performed much lower than native speakers with comparable C-Test scores when instructed to repeat the exact wording of sentences in their L2. Even though the C-Test was included as a covariate, the fact whether a participant was a native or a non-native speaker affected sentence recall performance. In contrast, only sentence recall performance accounted for variance in the C-Test scores.

One explanation for this finding is that sentence recall is a more fine-grained measure of language proficiency than the C-Test. In this case, sentence recall would uncover subtle deficits the C-test did not detect. A finding that can be interpreted as supporting this idea is the overall greater error rate in sentence recall as compared to C-test performance. Nonetheless, there was a considerable range for the C-Test scores both within the group of native speakers (75–99) and within the non-natives (81–99).

A potentially critical aspect concerns the use of written output in sentence recall, which might be more problematic for non-native than for native speakers. This output modality was chosen to maximize structural similarity to the C-Test. To ensure that the results reported here are not restricted to written output, we replicated the study with another small group of natives ($N = 9$) and near-natives ($N = 9$) using oral sentence recall. This study revealed similar results. Given these additional data and the fact that both tasks, sentence recall and the C-Test, required written output, it seems implausible that the relatively poor sentence recall performance for otherwise highly-proficient L2 speakers can be attributed to problems with the output modality. Furthermore, it has been suggested that written output reduces the demands for output control and thus the cognitive demands imposed by a recall task since it provides a written record of the already recalled words (Marsh et al., 2011). These considerations even suggest favoring written over oral recall when comparing natives' and near-natives' sentence recall performance.

A further caveat could be the fact that the group of native speakers was more heterogeneous with respect to their age and educational background than the non-native speakers were. In addition to a sample of university students, we included high school students and vocational school students. However, analogous analyses with a smaller sample in which only the L1 university students were included revealed similar results. The only difference was a descriptive advantage for L1 over L2 speakers on the C-Test, but with recall performance as a covariate [$F_{(1, 39)} = 34.75$; $p < 0.001$], the main effect for language group did again not reach significance ($F < 1$). The difference in age

and educational background thus influenced performance in the native sample but could not have caused the basic data pattern.

Assuming that the large discrepancy between sentence recall and C-Test performance is indeed caused by the need to process an L2, we now return to our underlying hypotheses. According to Potter and Lombardi (1990, 1998), sentence recall is a reconstruction based on the propositional structure generated during comprehension, plus lexical and syntactic priming. We assumed that under the instruction to recall the exact wording of a sentence, additionally, surface representations such as phonological ones would be maintained. This process of maintenance requires attention (Barrouillet et al., 2004). Sentence recall performance should thus decline if another process simultaneously draws on attention, as does second language comprehension and production. The fact that even highly proficient L2 speakers show poorer sentence recall performance than L1 speakers with the same mean C-Test scores supports this assumption. The present findings therefore highlight the fact that sentence recall is a task that requires substantial language skills as well as attention. Therefore, an explanation of how sentence recall works needs to consider both its linguistic and its working memory/attentional aspects.

As outlined above, it is plausible that the poor sentence recall performance of near-natives is due to a lack of automaticity of L2 (compared to L1) processing. The lesser degree of automaticity influences performance when the verbal task requires additional attentional control, as is the case with sentence repetition. In near-natives this difference in automaticity is not crucial for regular language processing and for successful education, as is indicated by the fact that the non-native speakers we tested performed native-like on the C-Test and had successful educational careers in their L2 German (all had acquired the German general qualification for university entrance). However, it remains indisputable that the degree to which language processing is automatized is also a criterion for language proficiency. It may be that—at the high end of L2 proficiency—the discrepancy between C-Test performance and sentence recall performance is a feasible way of measuring automaticity. Furthermore, the C-Test cannot fully capture the demands of online processing of transient information, as the to-be-completed texts are available for re-inspection throughout the task. Even though it is an established instrument for measuring language proficiency that correlates with more extensive test batteries (for an overview see Eckes, 2010), it is not a perfect measure of general language competence. There are certain verbal tasks that may require a similar degree of attentional control or pose similar maintenance demands as sentence recall does. Possible candidates are dual task situations (such as talking while driving, e.g., Becic et al., 2010), the resolution of long-distance anaphora (e.g., Daneman and Carpenter, 1980) or the processing of long-distance dependencies (e.g., Deane, 1991; Hawkins, 2004). With high attentional load in these tasks, sentence recall might predict L2 speakers' performance at least as well as the C-Test, even in mixed groups of natives and non-natives³. Further research that uses a larger battery of verbal tasks is required to address these questions.

³We thank one of the reviewers for this suggestion.

To have a strong test of our prediction that sentence recall underestimates L2 proficiency, we chose highly proficient L2 speakers. However, the same should hold true for a broader group of non-native speakers. As with increasing L2 proficiency automaticity goes up, sentence recall should also pose problems at lower levels of L2 proficiency. The implications of our findings are hence not restricted to the constrained group of near-natives. Still, a plausible exception to this exists: for speakers on the lower end of L2 proficiency, sentence recall could actually overestimate language skills. Since one can repeat lists of non-words, it is possible to base “sentence” recall on rote repetition and thus to repeat (parts) of sentences that one is not able to understand.

To conclude, our data suggest that sentence recall is not as good a measure as previously assumed when it is used as a predictor for how well a non-native speaker is able to communicate or to participate in education. This is particularly problematic when non-native and native speakers are analyzed jointly. When the goal is to identify differences within a group of non-natives or within a group of natives, sentence recall can still be an adequate measure. When sentence recall is used in language screenings for mixed groups (e.g., Fried and Briedigkeit, 2007), the problem might be addressed by gathering separate norms for native and non-native speakers.

ACKNOWLEDGMENTS

This research was supported by DFG grant Ru 891/6-1 to Ralf Rummer and Judith Schweppe. Thank you to Marie-Kristin Sommer for her help in running the experiment and to Anne Fürstenberg, Lena De Mol, and the two reviewers for helpful comments on an earlier draft.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fpsyg.2015.00063/abstract>

REFERENCES

- Aaronson, D., and Scarborough, H. S. (1976). Performance theories for sentence coding: some quantitative evidence. *J. Exp. Psychol.* 2, 56–70.
- Abutalebi, J. (2008). Neural aspects of second language representation and language control. *Acta Psychol.* 128, 466–478. doi: 10.1016/j.actpsy.2008.03.014
- Baddeley, A. D., Hitch, G. J., and Allen, R. J. (2009). Working memory and binding in sentence recall. *J. Mem. Lang.* 61, 438–456. doi: 10.1016/j.jml.2009.05.004
- Barrouillet, P., Bernardin, S., and Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *J. Exp. Psychol. Gen.* 133, 83–100. doi: 10.1037/0096-3445.133.1.83
- Barrouillet, P., and Camos, V. (2009). Interference: unique source of forgetting in working memory? *Trends Cogn. Sci.* 13, 145–146. doi: 10.1016/j.tics.2009.01.002
- Baur, R. S., and Meder, G. (1994). “C-Tests zur Ermittlung der globalen Sprachfähigkeit im Deutschen und in der Muttersprache bei ausländischen Schülern in der BRD [C-tests for the evaluation of general language skill in German and in the native language of migrant students in the FRG],” in *Der C-Test. Theoretische Grundlagen und praktische Anwendungen* [The C-test – theoretical basis and practical application], ed R. Grotjahn (Bochum: Brockmeyer), 151–178.
- Becic, E., Dell, G. S., Bock, K., Garnsey, S. M., Kubose, T., and Kramer, A. F. (2010). Driving impairs talking. *Psychon. Bull. Rev.* 17, 15–21. doi: 10.3758/PBR.17.1.15

- Birdsong, D. (2006). Age and second language acquisition and processing: a selective overview. *Lang. Learn.* 56, 9–49. doi: 10.1111/j.1467-9922.2006.00353.x
- Bosch, L., Costa, A., and Sebastián-Gallés, N. (2000). First and second language vowel perception in early bilinguals. *Eur. J. Cogn. Psychol.* 12, 189–221. doi: 10.1080/09541446.2000.10590222
- Brener, R. (1940). An experimental investigation of memory span. *J. Exp. Psychol.* 26, 467–482. doi: 10.1037/h0061096
- Clahsen, H., and Felser, C. (2006). How native-like is non-native language processing? *Trends Cogn. Sci.* 10, 564–570. doi: 10.1016/j.tics.2006.10.002
- Cowan, N. (1995). *Attention and Memory: An Integrated Framework*. Psychology Series, No. 26. Oxford; New York: Oxford University Press.
- Cowan, N. (1999). “An embedded-processes model of working memory,” in *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*, eds A. Miyake and P. Shah (New York, NY: Cambridge University Press), 62–101. doi: 10.1017/CBO9781139174909.006
- Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behav. Brain Sci.* 24, 87–114. doi: 10.1017/S0140525X01003922
- Daneman, M., and Carpenter, P. A. (1980). Individual differences in working memory and reading. *J. Verbal Learn. Verbal Behav.* 19, 450–466. doi: 10.1016/S0022-5371(80)90312-6
- Deane, P. D. (1991). Limits to attention: a cognitive theory of island constraints. *Cogn. Linguist.* 2, 1–63. doi: 10.1515/cogl.1991.2.1.1
- Diller, J., and Jordan-Diller, K. (2003). *Sentence Repetition Testing (SRT) and Language Shift Survey of the Tuki Language*. SIL Electronic Survey Reports, Vol. 10 (Dallas, TX: SIL International), 1–26.
- Eckes, T. (2010). “Der Online-Einstufungstest Deutsch als Fremdsprache (onDaF): theoretische Grundlagen, Konstruktion und Validierung [The online placement test ‘Deutsch als Fremdsprache (onDsF)’: theoretical foundations, construction, and validation],” in *Der C-Test: Beiträge aus der Aktuellen Forschung/The C-test: Contributions from Current Research*, ed R. Grotjahn (Hrsg.) (Frankfurt: Lang), 125–192.
- Eckes, T., and Grotjahn, R. (2006). A closer look at the construct validity of C-tests. *Lang. Test.* 23, 290–325. doi: 10.1191/0265532206lt330oa
- Fletcher, P. C., Shallice, T., and Dolan, R. J. (1998). The functional roles of prefrontal cortex in episodic memory. I. *Encoding Brain* 121, 1239–1248. doi: 10.1093/brain/121.7.1239
- Fried, L. (2008). Delfin 4: Diagnostik, Elternarbeit und Sprachförderung bei Vier-Jährigen in NRW. [Delfin 4: diagnosis, parental advice, and language promotion for 4 year olds in North Rhine-Westphalia.] *Schulverwaltung* 19, 300–302.
- Fried, L., and Briedigkeit, E. (2007). Delfin 4 – Hintergründe und Einblicke zum neuen System der Sprachstandsfeststellung und –förderung [Delfin 4 – background and insights in the new system for language proficiency assessment and promotion]. *Kompakt Spezial* 5, 10–11.
- Gauthier, S. V., and Madison, C. L. (1998). *Kindergarten Language Screening Test*. Austin, TX: PRO-ED.
- Grimm, H. (2001). *Sprachentwicklungstest für drei- bis fünfjährige Kinder. Diagnose von Sprachverarbeitungsfähigkeiten und auditiven Gedächtnisleistungen [A Test of Language Development for Three to Five Year Old Children. Diagnosis of Language Processing Skills and Auditory Memory Performance]*. Göttingen: Hogrefe.
- Grimm, H. (2003). *Sprachscreening für das Vorschulalter (SSV) [A Language Screening for Preschool Children]*. Göttingen: Hogrefe.
- Grotjahn, R. (2007). *The C-Test Bibliography: Electronic Version*. <http://www.c-test.de/deutsch/index.php?lang=de&content=bibliografie§ion=cetest#G> (Accessed 16 Feb, 2010)
- Hacker, W., Handrick, S., and Veres, T. (2002). *Lesespanntest*. [Reading span test.] Dresden: TU Dresden Eigenverlag.
- Hawkins, J. A. (2004). *Efficiency and Complexity in Grammars*. New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780199252695.001.0001
- Indefrey, P. (2006). A meta-analysis of hemodynamic studies on first and second language processing: which suggested differences can we trust and what do they mean? *Lang. Learn.* 56, 279–304. doi: 10.1111/j.1467-9922.2006.00365.x
- Lewandowsky, S., Oberauer, K., and Brown, G. D. A. (2009). No temporal decay in verbal short-term memory. *Trends Cogn. Sci.* 13, 120–126. doi: 10.1016/j.tics.2008.12.003
- Lombardi, L., and Potter, M. C. (1992). The regeneration of syntax in short term memory. *J. Mem. Lang.* 31, 713–733. doi: 10.1016/0749-596X(92)90036-W
- Marsh, J. E., Beaman, C. P., and Jones, D. M. (2011). “Source monitoring errors under conditions of distraction: The dependence of input modality on output mode,” *Paper presented at the 5th International Conference on Memory*, (York).
- Meijer, P. J. A., and Fox Tree, J. A. (2003). Building syntactic structures in speaking: a bilingual exploration. *Exp. Psychol.* 50, 184–195. doi: 10.1026//1617-3169.50.3.184
- Meyers, J. M., Volkert, K., and Diep, A. (2000). Sentence repetition test: updated norms and clinical utility. *Appl. Neuropsychol.* 7, 154–159. doi: 10.1207/S15324826AN0703_6
- Miller, E. K., and Cohen, J. D. (2001). An integrative theory of prefrontal cortex. *Annu. Rev. Neurosci.* 24, 167–202. doi: 10.1146/annurev.neuro.24.1.167
- Newcomer, P., and Hammill, D. (2008). *Test of Language Development–2 Primary*. Austin, TX: Pro-Ed.
- Perani, D., Abutalebi, J., Paulescu, E., Brambati, S., Scifo, P., Cappa, S. F., et al. (2003). The role of age of acquisition and language usage in early, high-proficient bilinguals: an fMRI study during verbal fluency. *Hum. Brain Mapp.* 19, 170–182. doi: 10.1002/hbm.10110
- Potter, M. C., and Lombardi, L. (1990). Regeneration in the short-term recall of sentences. *J. Mem. Lang.* 29, 633–654. doi: 10.1016/0749-596X(90)90042-X
- Potter, M. C., and Lombardi, L. (1998). Syntactic priming in immediate recall of sentences. *J. Mem. Lang.* 38, 265–282. doi: 10.1006/jmla.1997.2546
- Raatz, U., and Klein-Braley, C. (1982). “The C-Test - a modification of the cloze procedure.” in *Practice and Problems in Language Testing*, Vol. IV, eds T. Culhane, C. Klein-Braley and D. K. Stevenson (Colchester: University of Essex), 113–138.
- Radloff, C. F., and Hallberg, D. (1991). *Sentence Repetition Testing for Studies of Community Bilingualism*. Arlington, TX: Summer Institute of Linguistics.
- Rummer, R., and Engelkamp, J. (2001). Phonological information contributes to short-term recall of auditorily presented sentences. *J. Mem. Lang.* 45, 451–467. doi: 10.1006/jmla.2000.2788
- Rummer, R., and Engelkamp, J. (2003). Phonological information in immediate and delayed sentence recall. *Q. J. Exp. Psychol.* 56A, 83–95. doi: 10.1080/02724980244000279
- Rummer, R., Engelkamp, J., and Konieczny, L. (2003). The subordination effect: evidence from self-paced reading and recall. *Eur. J. Cogn. Psychol.* 15, 539–566. doi: 10.1080/09541440340000015
- Rummer, R., Schweppe, J., and Martin, R. C. (2013). Two modality effects in verbal short-term memory: evidence from sentence recall. *J. Cogn. Psychol.* 25, 231–247. doi: 10.1080/20445911.2013.769953
- Schweppe, J. (2006). *Shared Representations in Language Processing and Verbal Short-Term Memory: The Case of Grammatical Gender*. Saarbrücken: Dissertation at Saarland University. Available online at: <http://scidok.sulb.uni-saarland.de/volltexte/2006/851/>; URN: urn:nbn:de:bsz:291-scidok-8518
- Schweppe, J., and Rummer, R. (2007). Shared representations in language processing and verbal short-term memory: the case of grammatical gender. *J. Mem. Lang.* 56, 336–356. doi: 10.1016/j.jml.2006.03.005
- Schweppe, J., Rummer, R., Bormann, T., and Martin, R. C. (2011). Semantic and phonological information in sentence recall: experimental evidence and a neuropsychological single-case study. *Cogn. Neuropsychol.* 28, 521–545. doi: 10.1080/02643294.2012.689759
- Stowe, L. A., and Sabourin, L. (2005). Imaging the processing of a second language: effects of maturation and proficiency on the neural processes involved. *Int. Rev. Appl. Linguist. Lang. Teach.* 43, 329–353. doi: 10.1515/iral.2005.43.4.329
- Thompson-Schill, S. L., D’Esposito, M., Aguirre, G. K., and Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: a reevaluation. *Proc. Natl. Acad. Sci. U.S.A.* 94, 14792–14797.
- Wockenfuß, V., and Raatz, U. (2006). “Über den Zusammenhang zwischen Testleistung und Klassenstufe bei muttersprachlichen C-Tests [On the relation of test performance and class level in C-tests in the native language],” in *Der C-Test: Theorie, Empirie, Anwendungen - The C-Test: Theory, Empirical Research, Applications*, ed R. Grotjahn (Frankfurt/M: Lang), 211–242.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 25 September 2014; accepted: 13 January 2015; published online: 04 February 2015.

Citation: Schweppe J, Barth S, Ketzer-Nöltge A and Rummer R (2015) Does verbatim sentence recall underestimate the language competence of near-native speakers? *Front. Psychol.* 6:63. doi: 10.3389/fpsyg.2015.00063

This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology.

Copyright © 2015 Schweppe, Barth, Ketzer-Nöltge and Rummer. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.