

A Procedure for Improving the Effectiveness of a Cloze Test based on Classical True Score Theory Item Analysis Techniques

Researchers



Parviz Birjandi, Ph.D. in TEFL,
Islamic Azad University, Science & Research Campus,
E-mail: pbirjandi@yahoo.com
Parviz Panahi, M.A. in TEFL, Allame-e Tabatabai University
E-mail: parva-panahi@yahoo.com

چکیده

روایی و پایایی آزمون‌های «کلوز» (Cloze)، همواره یکی از دغدغه‌های آزمون‌سازان بوده است و این دو مقوله در پژوهش‌های بسیاری مورد بررسی قرار گرفته‌اند. در پژوهش حاضر که تلاشی است در جهت افزایش درصد روایی و پایایی آزمون‌های کلوز، از روش‌های آماری تحلیل نتایج آزمون برای انتخاب واحدهای آزمون استفاده شده است. در این تحقیق آزمون کلوز به‌عنوان آزمونی برای سنجش توانایی زبانی زبان‌آموزان در نظر گرفته شده و تلاش بر این بوده است که متن انتخاب شده و سطح زبانی آزمایش‌شوندگان ثابت بماند. نتایج بررسی نشان می‌دهد که روش به‌کار رفته در این پژوهش کارآمد بوده، چرا که توانسته است تا حد زیادی روایی، پایایی و پراکندگی نمرات را بهبود بخشد.

کلید واژه‌ها: روایی، پایایی، نظریه‌ی سنتی ارزش‌یابی

Abstract

Cloze reliability and validity have been a concern for language testers who have addressed the issues in many studies on cloze testing. The present study is an attempt to improve the reliability and validity of a cloze test through classical item analysis techniques as criteria for item selection. Examining cloze as a measure of second language proficiency, this study describes results obtained from altering item selection method of the cloze procedure while holding text and student ability constant. The results of the study indicate the method applied in the study namely the "tailored method" enjoys certain advantages which make its application more justified. The dispersion of scores, and the reliability and validity of the test results of the study showed that the "tailored method" was effective for its purpose. In all respects, the use of item analysis techniques to improve the cloze test constructed in the study for norm-referenced purposes seems to have succeeded, especially in the light of the relatively narrow range of language abilities represented in the sample.

Key Words: reliability, validity, classical true score measurement theory

really the case, we should not expect to observe (significant) differences between the planned and unplanned performance of the learners as far as articles are concerned. Hence the logical position seems to be to investigate the role of planning in relation to different components of language and subcomponents of language.

Conclusion

Research on pre-task planning indicates that L2/FL learners tend to use more varied and complex language in planned conditions and this upgraded performance would contribute to interlanguage development. This paper has attempted to shed light on our understanding of planning and its effects on L2/FL performance. It would appear from the results that whether or not strategic planning has any effect on accuracy depends on a variety of factors such as the extent to which particular learners are oriented towards accuracy, learners' strategies, the type of task, as well as the linguistic feature under investigation. Clearly, more research is needed to identify how planning interacts with task design variables, implementational procedures and learner factors.

References

- Crookes, G. (1989). Planning and Interlanguage Variation. *Studies in Second Language Acquisition*, 11, 367-383.
- Ellis, R. (1987). Interlanguage Variability in Narrative Discourse: Style in the Use of the Past Tense. *Studies in Second Language Acquisition*, 9, 12-20.
- Foster, P., & Skehan, P. (1996). The influence of Planning and Task Type on Second Language Performance. *Studies in Second Language Acquisition*, 18, 299-323.
- Foster, P., & Skehan, P. (1999). The Influence of Source of Planning and Focus of Planning on Task-Based Performance. *Language Teaching Research*, 3, 215-247.
- Hulstijn, J. H. and Hulstijn, W. (1984). Grammatical Errors as a Function of Processing Constraints and Explicit Knowledge. *Language Learning*, 34, 23-43.
- Iwashita, N., Elder, C., and McNamara, T. (2001). Can We Predict Task Difficulty in an Oral Proficiency Test? Exploring the Potential of an Information-Processing Approach to Task Design. *Language Learning*, 51, 401-36.
- Master, P. (1990). Teaching the English Articles as a Binary System. *TESOL Quarterly*, 24, 461-478.
- Mehnert, U. (1998). The Effect of Different Lengths of Time for Planning on Second Language Performance. *Studies in Second Language Acquisition*, 20, 52-83.
- Mitchell, R. and F. Myles. 2004. *Second Language Learning Theories*. 2nd ed. Oxford: Oxford University Press.
- Ortega, L. (1999). Planning and Focus on Form in L2 Oral Performance. *Studies in Second Language Acquisition*, 21, 108-148.
- Parrish, B. (1987). A New Look at Methodologies in the Study of Article Acquisition for Learners of ESL. *Language Learning*, 37, 361-383.
- Pica, T. (1984). Methods of Morpheme Quantification: Their Effect on the Interpretation of Second Language Data. *Studies in Second Language Acquisition*, 6, 69-78.
- Skehan, P. (1996). A Framework for the Implementation of Task-Based Instruction. *Applied Linguistics*, 17, 38-62.
- Skehan, P., & Foster, P. (1997). Task Type and Task Processing Conditions as Influences on Foreign Language Performance. *Language Teaching Research*, 1(3), 1-27.
- Wigglesworth, G. (1997). An Investigation of Planning Time and Proficiency Level on Oral Test Discourse. *Language Testing*, 14, 85-106.
- Willis, J. (1996). *A frame Work for Task-Based Learning*. London: Longman.
- Yuan, F., & Ellis, R. (2003). The effects of Pre-Task Planning and On-Line Planning on Fluency, Complexity and Accuracy in L2 Monologic Oral Production. *Applied Linguistics*, 24, 1-27.

of planning time. Actually, if we convert the numbers to percentages, the students in the first group (condition 1), who had five minutes of planning time used definite and indefinite articles accurately in 66% and 88% of the times respectively. Similarly, the students in the second group (condition 2) who had no planning time used articles *the* and *a/an* accurately in 71% and 81% of the time. As it is evident, the difference between the two groups regarding the accurate use of articles is quite negligible.

What is interesting in the data is that the second group used articles *the* and *a/an* approximately two times more than the first group. The explanation may be found in the amount of time spent for writing the texts by the two groups of students. The students of group one had five minutes planning time and after that time they were asked to write their texts in 10 minutes. In contrast, the students of group two didn't have any planning time and were asked to start writing after seeing the comic strip immediately and finish in 15 minutes. Since the second group of students had more time to write, that may be the reason for their more use of articles.

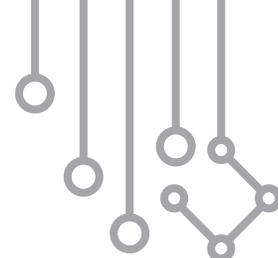
Table 2. Accuracy Rate of Article Use in Planned and Unplanned Written Performance

Use of <i>the</i> & <i>a/an</i>	Accurate use of <i>the</i>	Inaccurate use of <i>the</i>	Accurate use of <i>a/an</i>	Inaccurate use of <i>a/an</i>	Total number of <i>the</i>	Total number of <i>a/an</i>
Condition 1	20	10	15	2	30	17
Condition 2	45	18	25	6	63	31

Discussion

The present study intended to investigate the role of planning in

learners' accurate performance on two morphosyntactic devices, i.e., definite and indefinite articles. The results showed no significant difference between the group which was given time to plan before starting to write and the group which was not allotted the time and had to start the task immediately after the instruction. What does this imply about the role of planning? This study seems to support the findings of studies carried out by such researchers as Crookes (1989), Yuan and Ellis (2003), and Elder, and McNamara (2001) which assume no or little role for planning in learners' linguistic performance. However, in light of evidence to the contrary Ellis (1987), Foster and Skehan (1996) Skehan and Foster (1997), Wigglesworth (1997) among others, it would be an oversimplification to view this issue (as well as all other issues in language learning), as a matter of all or none and encourage more research to support either side. The fact of the matter is that, different linguistic devices are of different syntactic and/or semantic complexity levels, saliencies, and play differential roles in communicating meaning. For instance, there is evidence to show that due to their semantic complexity, articles are notoriously difficult to internalize and produce. Besides, they carry little semantic load (Mitchell and Myles, 2004) in comparison to content words and might not attract the learners' attention and remain rather unattended even if learners are spared a few minutes for planning and preparation. If this is



It was assumed that 5 minutes of pre-task planning would benefit the first group since planning was assumed to render the completion of the task cognitively less demanding. On the contrary, the same task would be more difficult for the second group and required heavier cognitive demand (Table 1).

Table 1. Design of the task

Dimension	Requirements of dimension	predicted difficulty
+ Planning time	5 minutes for looking at the pictures	less difficult (lighter cognitive demand)
- Planning time	no time for looking at the pictures	more difficult (heavier cognitive demand)

In order to analyze the participants' written language performance in terms of its accuracy and planning time, the researchers collected written texts from the participants. The data consisted of the narrations of a wordless comic strip in the written form under two conditions of planning time.

The participants' use of articles was the focus of the study, because based on their teaching experience, the researchers found that Iranian learners of English usually struggle with these linguistic items more than other items in their writings. Our use of a specific measure of accuracy derives from Crookes (1989) and Ellis (1987), who looked for specific linguistic features in their data (e.g., third person -s and past tense markings respectively).

The researchers counted every occurrence of article in the participants' written

performance and noted whether they had been used correctly or incorrectly based on the kind of noun phrase (NP) it was preceding, i.e., whether the NP was proper or common, specific or generic, given or new, noncount or count, and/or plural or singular (see Master, 1990; Parrish, 1987; and Pica, 1984, for examples of studies performed on article acquisition and method of analyzing article usage). Moreover, it should be noted that the calculation of inaccurate versus accurate language production of the participants' writings was conveyed only by their use of *a/an* and *the* and lack of these articles in the participants' written texts was not taken into account because the amount of this kind of articles was negligible in the participants' writings. In order to obtain the inter-rater reliability, the data were also analyzed by the teacher of both of the groups of students and an inter-rater reliability quotient of 1.00 was obtained that indicated the consistency with which students' performances were evaluated.

Results

Before analyzing the data, the researchers hypothesized that participants' accuracy in writing would increase based on the amount of planning time they were given. As a result, the researchers postulated that the "unplanned" texts would be less accurate.

The analysis of the data (Table 2) showed that students from both conditions 1 and 2 accurately used *a/an* and *the* most of the time in their written tasks despite the amount





using a general measure of accuracy, found that three minutes of planning time had no effect on accuracy in a testing situation. There are reports of mixed findings as well. In Ortega's (1999) study, planning led to greater accuracy in the case of Spanish noun-modifier agreement but not in the case of articles. Mehnert (1998) also reports a significant difference in the accuracy of one-minute planners over non-planners. However, the five-minute and ten-minute planners performed at the same overall level of accuracy as the one-minute planners. And finally, Foster and Skehan's (1999) study of the effects of the source of planning found that accuracy was greatest when the planning was teacher-led. However, rather surprisingly, directing learners' attention to form as opposed to content during planning had no effect on accuracy.

Considering the above controversy, the present study is an attempt to investigate the relationship between task characterization and task performance. More specifically, it intends to investigate the relationship between task performance condition and response accuracy. In order to achieve the specific objective of the study, the following question has been addressed:

What is the effect of pre-task planning on the accurate use of articles "a/an" and "the" in narrative tasks performed by students of the same proficiency level?

It is hypothesized that access to forms (in our case, target-like use of articles "a/an" and "the") that have not been fully automatized

would benefit learners if they are allowed planning time before performance.

Method

In this study, 12 intermediate students of a private EFL teaching institute in Mashad participated. All of the participants were males and they had been studying English as a foreign language at this private institute for the last two years. The mean age of the participants was 15 and all of them were speakers of Persian. In order to determine the proficiency level of the participants, a standard TOEFL test was administered at the institute. Then, based on the results, 12 participants were randomly chosen and assigned to two equal groups, each consisting of six members. The participants were presented with a wordless comic strip which prompted them to narrate under different processing conditions. Planning time was used as a variable to generate two levels of task performance. In Condition 1, the task was kept simple by providing learners (group 1) with 5 minutes of pre-task planning time and students were asked to narrate the story in the written form in the present tense within 10 minutes while looking at the comic strip. In Condition 2, learners (group 2) were asked to proceed in the same way but this time no pre-task planning time was allowed so they had to narrate the story almost immediately after seeing the strip and under 15 minutes deadline. One measure of accuracy (the target-like use of articles) was used to code the narrated stories.

Differences are also observed in learners' performance depending on whether the communicative task is spoken or written. Foreign language learners tend to commit relatively more errors in spontaneous speech than in written discourse.

Although committing errors is inevitable in the process of learning a language, we need to establish criteria for detecting errors. The measurement of language production has long proved problematic for researchers. To a large extent, measures of production have been chosen based on intuition or data, rather than theory. An exception is Skehan. Skehan (1996) distinguishes between fluency, accuracy, and complexity, drawing on his theoretical claims about a dual-competence system and trade-offs in learners' focus of attention.

The following review of the research will discuss the results obtained by different studies in relation to the accuracy aspect of production (the ability to avoid error in performance, possibly reflecting higher levels of control in the language). Accuracy has been defined differently in different studies, among which only "target-like use of articles" is adopted in the present study.

Background

Numerous studies have been carried out on the effects of planning on learners' production (e.g., Crookes, 1989; Ellis, 1987; Mehnert, 1998; Ortega, 1999; Wigglesworth, 1997). These studies in general hypothesized that planning time would enable L2/FL learners to produce

more accurate, higher levels of lexical complexity and fluent speech. However, results are rather mixed, and this is particularly true for accuracy. While some studies showed more accurate production under the planned condition (e.g., Ortega, 1999), other studies found that this increased accuracy was evident only for certain tasks (e.g., Foster and Skehan, 1996; Mehnert, 1998) and on certain measures (e.g., Crookes, 1989).

Controversy seems to be greater in the case of the effects of strategic planning on accuracy. Ellis (1987), for instance, found that providing opportunities for strategic planning resulted in more accurate use of the regular past tense. Foster and Skehan (1996) also report that both undetailed and detailed planners produced fewer errors than the non-planners on the decision-making task, and that the undetailed planners were only more accurate than the non-planners on the personal task, while no effect for planning on accuracy was evident on the narrative task. Besides, Skehan and Foster (1997) found that planning (undetailed) led to greater accuracy on the personal and narrative tasks but not on the decision-making task. Wigglesworth (1997), too, reports that her high proficiency learners benefited from planning the most difficult task where verb morphology and articles were concerned but found no effect for plurals. On the other hand, Crookes (1989) and Yuan and Ellis (2003), found that planning had little effect on accuracy. In a similar vein, Iwashita, Elder, and McNamara (2001),

Abstract

Research has so far shown that planning allows for greater use of metalinguistic knowledge and, hence, results in fewer errors as far as oral production is concerned. The present study is an attempt to investigate the potential of pre-task planning in developing second/ foreign language learners' ability in writing. More specifically, it examines the written data produced by Iranian learners of English under two conditions, i.e., planned and unplanned, focusing exclusively on their degree of accuracy in using definite and indefinite articles. The participants in the study were provided with a wordless comic strip on the basis of which they had to narrate a story under two processing conditions. Target-like use of articles was used as the measure of accuracy in coding the narrated stories. The findings of the study revealed that, first of all, Iranian students tend to use indefinite articles *a/an* more accurately than definite article *the* regardless of planning. Furthermore, planning has a more noticeable effect on the accurate production of indefinite articles *a/an* than definite article *the*. Hence, it can be suggested that, though planning cannot categorically affect the accuracy of linguistic items, it can be recommended because it can provide learners with an opportunity to scrutinize their own interlanguage system and to control their linguistic knowledge in communicative conditions.

Key Words: pre-task planning, writing, accuracy, indefinite article, definite article.

Introduction

If learners learning another language are given time to plan their discourse, written or spoken, then should their L2/ FL production contain fewer errors and be more linguistically accurate than if they weren't given any planning time? The initial response would probably be "yes", but does that really occur? Does the type of task and modality of language use (written vs. spoken) also play a role? These questions form the basis of this study that will look at the written texts (both planned and unplanned) of Iranian learners of English, focusing mainly on their accurate use of the English article system under both conditions.

The decision to focus on planning time and its effect on written discourse came about for one main reason. Much of the research carried out on planning time and second language performance has dealt with learners' oral production as measured by their fluency, complexity,

and accuracy in the target language (Crookes, 1989; Foster and Skehan, 1996; Mehnert, 1998; Wigglesworth, 1997). Studies conducted solely on planning time and written production aren't as prevalent in the literature, and when they do occur, they are typically part of a larger study investigating planning time with both written and spoken discourse (Ellis, 1987). In order to examine the influence of planning time on the written discourse of L2/FL learners, the researchers chose to focus only on the learners' accuracy in producing definite and indefinite articles.

Compared to spontaneous speech, planned discourse allows for greater use of metalinguistic knowledge and results in fewer errors. Time seems to play a determining role. Hulstijn and Hulstijn (1984) suggest that poor learners need more time to produce speech material because they have little control over their linguistic awareness.