Tabaran Institute of Higher Education ISSN 2476-5880
International Journal of Language Testing
Special Issue, 2023: *Advanced Psychometric Methods in Language Testing*
Research Note

Analysis of C-Tests with the Equidistance and the Dispersion Models

Ziad Ahmad Alabdallah[1], Ismail Abdulwahhab Ismail[2], Hamid Khalaf Mutar[3], Aisha Ahmed[4], Tawfeeq Alghazali[5], Mohammad Salman Mansoor[6], Akram Ali Anber[7], Yusra Mohammed Ali[8], Maabreh Hatem Ghaleb[9], Goncharova Velichka Georgievna[9]

Abstract
C-tests are commonly used as measures of second-language reading comprehension and general language proficiency. Analysis of C-tests with item response theory models is problematic due to the interdependent structure of C-test items or gaps. An approach to facilitate item response theory (IRT) analysis of C-tests involves treating each passage as a polytomous super-item. This approach facilitates the application of polytomous IRT models for ordered response data to the C-test passages. Usually, Andrich's rating scale model or Masters (1982) partial credit model are used to analyse the data. In this study, we aim to employ two alternative modelling techniques, namely, the equidistant model (Andrich, 1982) and the dispersion model (Rost, 1988) to C-test data. Our findings showed that the C-test data have a good fit to both models and can be used for psychometric analysis of C-test passages. Information criteria showed that the dispersion model has a better fit compared to the equidistance model.

Keywords: C-Test, dispersion model, equidistance model, local item dependence, polytomous Rasch models.

1. Introduction

A C-Test is a language proficiency test that measures a person's ability to comprehend and use a language. It is a type of language test that assesses all four language skills of reading, listening, writing, and speaking (Klein-Braley, 1997). The test consists of 4-8 short independent texts in which the second half of every second word is removed. The first and the last sentences in each passage remain intact to provide some context for text comprehension. The task of the test-

[1] College of Veterinary Medicine, Al-Furat University, Deirez-Zor, Syria
[2] English Department, Al-Noor University College, Nineveh, Iraq
[3] English Language and Literature Department, Al-Mustaqbal University College, Babylon, Iraq
[4] College of Pharmacy, Al-Farahidi University, Baghdad , Iraq
[5] College of Media, Department of journalism, The Islamic University in Najaf, Najaf, Iraq
[6] Department of Media, Mazaya University College, Iraq
[7] Al-Esraa University College, Baghdad, Iraq
[8] Department of Medical Laboratory Technics, Al-Zahrawi University College, Karbala, Iraq
[9] Peoples' Friendship University of Russia, Moscow, Russia

taker is to fill in the missing letters in the text. The difficulty level of a C-test increases as the length of the text increases, making it an effective tool for measuring language proficiency at different levels (Raatz & Klein-Braley, 2002).

The C-Test is actually a variation of the more familiar cloze test. The main difference between a C-Test and a cloze test is that in a cloze test whole words are deleted and examinees have to fill in the missing words but in a C-Test half of the words are deleted which means that examinees have to supply the missing letters. Another major difference between the two tests is that the rate and start point of deletions in cloze tests are not fixed and can vary from 7 to 10 but the rate of and the point of start of deletions in C-tests is fixed (it is always 2) and starts from the second word in the second sentence. The first sentence of the text serves as a context for the rest of the passage, which contains gaps that need to be filled with the correct letters. The difficulty level of the test can be adjusted by varying the length of the text, the number of gaps, and the complexity of the vocabulary used (Klein-Braley, & Raatz, 1984). Overall, constructing a C-Test requires careful attention to detail and an understanding of language proficiency testing principles.

C-tests have been found to have high validity in measuring language proficiency in various languages, including English, German, and French. The validity evidence for C-tests includes:

1. Content validity: C-tests are designed to measure language proficiency by testing the ability to comprehend and produce grammatically correct sentences. The content of the test is based on a representative sample of language use (Klein-Braley, 1997).
2. Construct validity: C-tests have been found to correlate highly with other measures of language proficiency, such as standardized tests like TOEFL or CEFR. C-tests have been shown to predict success in academic or professional settings where language proficiency is required. C-tests have been found to correlate with other measures of cognitive abilities, such as working memory and attention. Overall, c-tests are considered a reliable and valid measure of language proficiency in various contexts (Rasoli, 2021; Sigott, 2004).

1.1. Local item dependence in C-tests

The problem of local item dependence (LID) refers to the situation where the responses to one item in a test are influenced by the responses to another item in the same test. This can lead to inflated estimates of reliability and validity, as well as biased estimates of item difficulty and discrimination (Baghaei & Ravand, 2019). It can also affect the interpretation of test scores and make it difficult to compare scores across different tests or populations. To address this problem, researchers may use item response theory models that account for local item dependence (Baghaei & Ravand, 2016).

Because items in C-test passages are in fact the gaps and these gaps occur within the same passage and are very close to each other, the mere structure of the C-test is bound to produce LID. To account for local item dependence in C-tests, several methods can be used:

1. Item analysis: Conducting an item analysis can help identify items that are locally dependent. This involves examining the residual correlation between each pair of items and removing items that have high residual correlations (Christensen, 2017).
2. Item bundles: Item bundles involve grouping items into smaller subsets or super-items based on their shared prompts. That is, items which share the same prompt are parceled together and a testlet is made. This way, the dependency between the gaps is removed because the unit of analysis becomes the testlet. Polytomous IRT models are employed for analysis (Forthmann et al, 2020).
3. Testlet response theory: this is an IRT model which accounts for LID (Alpizar et al., 2023; Eckes & Baghaei, 2015).
4. Loglinear Rasch Model: In this method locally dependent items are identified and are allowed to correlate (Baghaei & Christensen, 2023).

The most commonly used method of accounting for LID in C-tests is the item bundle approach. As explained above, in the item bundle approach, items or gaps which are within the same passage are aggregated and a polytomous item is constructed. That is, a C-test with four passages (each passage with 20 gaps) becomes a test with four polytomous items (each item having 21 categories). Then a polytomous Rasch or IRT model is applied on the passages. Item bundles are also called testlets and super-items (Dhyaaldian et al., 2022a/2022b; Eckes, 2011; Hussein et al., 2022; Syman, 2023).

Previous research had showed that for the analysis of item bundles in C-tests, mostly the rating scale model (RSM; Andrcih, 1978) and Masters' partial credit model (PCM; Masters, 1982) have been used. There are several studies which show the applicability of these two polytomous Rasch models to C-Test data (Baghaei, 2021; Eckes, 2011). There are also two studies which indicate the suitability of the continuous Rasch model (CRM; Müller, 1987) for C-test analysis (Arras et al., 2002; Eckes, 2011).

The purpose of the current study is to apply two other polytomous Rasch models (Rasch, 1960/1980) to C-tests, namely, the equidistant model (Andrich, 1982) and the dispersion model (Rost, 1988). These two models are different parameterizations of the PCM (Master's 1982). In the equidistance model, the distance between the categories of the Likert items is assumed to be the same in each item. This model was specifically suggested for modeling LID (Baghaei, 2010). The dispersion model is a combination of the RSM and the equidistance model. In the dispersion model (like the PCM) each item has a separate set of threshold parameters but unlike the equidistance model, the distances between them are not equal. However, it is not as unrestricted as the PCM, since the number of response categories in each item should be the same. The advantage of these special cases compared to the PCM is that they are more parsimonious and fewer parameters are estimated (von Davier, 2014).

2. Method

2.1. Instrument and Participants

A C-test battery containing five independent passages was analyzed in this study. Each passage, which was 111 to 138 words in length, contained 20 gaps. The first and the last sentences in each passage were intact. A sample of 186 Iraqi undergraduate students (105 females and 81 males) of English as a foreign language at Al-Noor University College, Nineveh, Iraq took the C-test. The age range was from 19 to 39 (M=22.87, SD=3.91). The test was given as part of their final exam in reading comprehension in English as a foreign language.

2.2. Analyses and Results

The data were analyzed with the equidistance model of Andrich (1982) and the dispersion model of Rost (1988). The WINMIRA software program (von Davier, 2000) was used for model estimation. Since WINMIRA does not analyze polytomous items with more than 9 categories, the response categories in each C-test super item (i.e. passage) were collapsed and 7 categories were constructed for each item by merging every three adjacent categories.

Table 1 shows item locations, dispersion, and threshold parameters (τ) for each item in the equidistance model. As the threshold values show the distances between the thresholds are the same within each item but not across items. The dispersion parameter is half the threshold distance. The smaller the threshold distances, the more discriminating the items are (Andrich, 1982).

Table 1.
Location and Threshold Parameters for the Equidistance Model

Item	Location	Dispersion	τ1	τ2	τ3	τ4	τ5	τ6
1	.072	-.026	.204	.152	.099	.046	-.007	-.059
2	-.134	.006	-.168	-.155	-.142	-.128	-.115	-.102
3	-.486	.034	-.659	-.590	-.521	-.452	-.383	-.314
4	-.091	.018	-.158	-.148	-.110	-.073	-.036	.002
5	-.042	-.043	.175	.088	.001	-.086	-.173	-.260

Table 2 shows item locations, dispersion, and threshold parameters (τ) for each items in the dispersion model. As the threshold values show the distances between the thresholds are not the same within and across items.

Table 2.
Location and Threshold Parameters for the Dispersion Model

Item	Location	Dispersion	τ1	τ2	τ3	τ4	τ5	τ6
1	.282	.025	.625	.448	-.162	-.422	-.479	1.687
2	.072	.013	.473	.273	-.361	-.644	-.724	1.419
3	-.244	-.052	.488	.155	-.611	-1.028	-1.241	.769
4	.130	.043	.383	.242	-.333	-.557	-.578	1.624
5	.135	-.029	.753	.466	-.245	-.624	-.791	1.265

Table 3 reports the information criteria for the two models. All the three criteria AIC, BIC, and CAIC unanimously indicate that the dispersion model has a better fit than the equidistance model.

Table 3.
Information Criteria for the Equidistance and the Dispersion Model

Model	AIC	BIC	CAIC
Equidistance	3491.30	3523.56	3533.56
Dispersion	3235.94	3281.10	3295.10

3. Discussion and Conclusion

Polytomous Rasch models help in accounting for local item dependence by allowing for the estimation of item parameters that account for the relationship between items within a test or scale. Local item dependence occurs when two or more items in a test are related to each other, beyond what can be explained by their relationship with the underlying construct being measured. This can lead to biased estimates of person ability and item difficulty, as well as reduced measurement precision.

To account for LID in C-Tests, mostly Andrich's rating scale model has been used. In this study, two polytomous Rasch models, hitherto new for C-test analysis, were employed for modeling LID in C-tests. Responses of 186 students of English as a foreign language to four C-test passages were analyzed by considering each passage as a testlet. The equidistance model (Andrich, 1982) and the dispersion model (Rost, 1982) were then applied to the four C-test passages separately. Our findings showed that while both models work well for C-test analysis, the dispersion model of Rost (1988) has a better fit to C-test data. These two models can be added to the toolkit of language testers, applied psychometricians, and test developers for language test analysis and Likert-type psychological items.

References

Alpizar, D., Li, T., Norris, J. M., & Gu, L. (2023). Psychometric approaches to analyzing C-tests. *Language Testing*, *40*(1), 107–132. DOI: 10.1177/02655322211062138

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, *43*, 561–573. DOI: 10.1007/BF02293814

Andrich, D. (1982). An extension of the Rasch model for ratings providing both location and dispersion parameters. *Psychometrika*, *47*, 105–113. DOI: 10.1007/BF02293856

Arras, U., Eckes, T., & Grotjahn, R. (2002). C-Tests im Rahmen des "Test Deutsch als Fremdsprache" (TestDaF): Erste Forschungsergebnisse. In Grotjahn, R. (Ed.), Der C-Test: Theoretische Grundlagen und praktische Anwendungen (Vol. 4, pp. 175-209). Bochum: AKS-Verlag.

Baghaei, P., & Christensen, K. B. (2023). Modelling local item dependence in C-Tests with the Loglinear Rasch Model. *Language Testing*, Advanced online publication, DOI: 10.1177/02655322231155109

Baghaei, P. (2021). *Mokken scale analysis in language assessment*. Muenster: Waxmann.

Baghaei, P., & Ravand, H. (2019). Method bias in cloze tests as reading comprehension measures. *Sage Open, 9*, 1-8. DOI: 10.1177/2158244019832706

Baghaei, P., & Ravand, H. (2016). Modeling local item dependence in cloze and reading comprehension test items using testlet response theory. *Psicológica*, *37*, 85-104.

Dhyaaldian, S. M. A., Al-Zubaidi, S. H., Mutlak, D. A., Neamah, N. R., Albeer, M. A., Hamad, D. A., Al Hasani, S. F., Jaber, M. M., & Maabreh, H. G. (2022). Psychometric evaluation of cloze tests with the Rasch model. *International Journal of Language Testing*, *12*(2), 95-106. DOI: 10.22034/IJLT.2022.157127

Dhyaaldian, S. M. A., Kadhim, Q. K., Mutlak, D. A., Neamah, N. R., Kareem, Z. H., Hamad, D.A., Tuama, J. H., Qasim, M. S. (in press). A comparison of polytomous Rasch models for the analysis of C-Tests. *International Journal of Language Testing*, *12*(2), 107-117. DOI: 10.22034/IJLT.2022.157128

Eckes, T. (2011). Item banking for C-tests: A polytomous Rasch modeling approach. *Psychological Test and Assessment Modeling*, *53*(4), 414-439.

Eckes, T., & Baghaei, P. (2015). Using testlet response theory to examine local dependence in C-tests. *Applied Measurement in Education*, *28*(2), 85-98. DOI: 10.1080/08957347.2014.1002919

Forthmann, B., Grotjahn, R., Doebler, P., & Baghaei, P. (2020). A comparison of different item response theory models for scaling speeded C-Tests. *Journal of Psychoeducational Assessment, 38*, 692-705. DOI: 10.1177/0734282919889262

Hussein, R. A., Sabit, S. H., Alwan, M. G., Wafqan, H. M., Baqer, A. A., Ali, M. H., Hachim, S. K., Sahi, Z. T., AlSalami, H. T., & Sulaiman, B. F. (2022). Psychometric evaluation of dictations with the Rasch model. *International Journal of Language Testing*, *12*, 118–127. DOI: 10.22034/IJLT.2022.157129

Klein-Braley, C. (1997). C-tests in the context of reduced redundancy testing: An appraisal. *Language Testing*, *14*(1), 47-84. DOI: 10.1177/026553229701400104

Klein-Braley, C., & Raatz, U. (1984). A survey of research on the C-test. *Language Testing*, *1*(2), 134-146. DOI: 10.1177/026553228400100202

Masters, G. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*, 147–174. DOI: 10.1007/BF02296272

Müller, H. (1987). A Rasch model for continuous ratings. *Psychometrika*, *52*, 165–181. DOI: 10.1007/BF02294232

Raatz, U., & Klein-Braley, C. (2002). Introduction to language testing and to C-Tests. In J. A. Coleman, R. Grotjahn, & U. Raatz (Eds.), *University language testing and the C-test* (pp. 75–91). Bochum: AKS-Verlag.

Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests* (expanded Ed.). Chicago, IL: University of Chicago Press.

Rasoli, M. K. (2021). Validation of C-test among Afghan students of English as a foreign language. *International Journal of Language Testing, 11*(2), 109-121.

Rost, J. (1988). Measuring attitudes with a threshold model drawing on a traditional scaling concept. *Applied Psychological Measurement, 12*, 397–409. DOI: 10.1177/014662168801200408

Syman, K., Alallo, H. M. I., Mohammed, A., Hassan, A. Y., Suleiman, O. W., Ali, Y. M., Ghaleb, M. H., Mikhailovna, K. A., John, E. A., & Georgievna, G. V. (2023). Psychometric modelling of reading aloud with the Rasch model. *International Journal of Language Testing, (Special Issue)*, 62-68. DOI: 10.22034/IJLT.2023.386577.1235

von Davier, M. (2014). Rasch polytomous models. In Michalos, A. C. (Ed.) *Encyclopedia of quality of life and well-being research* (pp. 5396-5399). Springer, Dordrecht. DOI: 10.1007/978-94-007-0753-5_2412

von Davier, M. (2000). *WINMIRA 2001 – Software Manual*. IPN: Institute for Science Education. Student Version available at: http://www.von-davier.com