Validity and Reliability of the Inductive Reasoning Test for Children – IRTC

Validade e Precisão do Teste de Raciocínio Indutivo para Crianças – TRIC

Monalisa Muniz[*, a], Alessandra Gotuzo Seabra[b] & Ricardo Primi[c]

[a]*Universidade do Vale do Sapucaí-UNIVÁS, Pouso Alegre, Brasil,*
[b]*Universidade Presbiteriana Mackenzie, São Paulo, Brasil &* [c]*Universidade São Francisco, Itatiba, Brasil*

Abstract

This study aimed to search for validity and reliability evidences for 'A and B Forms' of the Inductive Reasoning Test for Children – IRTC. A total of 417 students, both sexes, from 1st to 5th grade, between 6 and 11 years old participated of the study. 'A Form' was administered to 219 children and 'B Form' to 198. The investigation of the internal structure of both forms indicated that the correlations between the items may be explained by a single factor which is understood as inductive reasoning. Appropriate reliability by internal consistency was found, except for the 1st and 2nd grades, and regarding the latter, only on 'B Form'. The results also revealed the sensitivity of IRTC in showing differences in the performance of students attending early versus late grade levels. Despite the favorable data, after the reformulation of some items more research is needed, so that 'A and B Forms' are able to discriminate the performance of children in a greater number of grades.
Keywords: Intelligence, psychological test, psychometrics.

Resumo

Esse estudo buscou evidências de validade e precisão para as Formas A e B do Teste de Raciocínio Indutivo para Crianças – TRIC. Participaram 417 alunos de 1ª a 5ª série, ambos os sexos e com idades entre 6 e 11 anos. A Forma A foi respondida por 219 crianças e a Forma B por 198. A investigação da estrutura interna das duas formas apontou que as correlações entre os itens podem ser explicadas por um único fator entendido como o raciocínio indutivo. As precisões por consistência interna encontradas foram adequadas, exceto para a primeira série e para a Forma B na segunda série. Os resultados também revelaram sensibilidade das formas do TRIC para diferenciar o desempenho entre as séries iniciais e finais. Apesar dos dados favoráveis, ainda são necessárias mais pesquisas para que as Formas A e B consigam discriminar o desempenho das crianças em um maior número de séries.
Palavras-chave: Inteligência, teste psicológico, psicometria.

The construction of a psychological instrument is based on the concepts of reliability, validity, standardization and normalization, because it is only with all these factors aggregated that it is possible to consider a test appropriate for use. Moreover, according to Urbina (2006), when creating a test, it is essential to balance the objective adequate theoretical basis. The theory that forms the basis for a test is the first step because, in fact, it is necessary to know the construct that the test intends to evaluate and, then, to support this construct theoretically, in order to know it and understand it, that allows the elaboration of appropriate items for the test, as well as ways of scoring and interpreting of results.

However, according to Wilhelm (2005), this does not always occur in the construction of tests to assess reasoning. According to the author, the construction of such instruments is usually approached by applying psychometric criteria, that is, the items are constructed and applied to a sample of subjects to, subsequently, perform analyses of psychometric parameters, such as correlation and factor analysis. From this a conclusion on whether the test is appropriate or not is drawn, often without examining the psychological rational for the data. In fact, the construction of a reasoning test should be guided by indicators based on theories, derived from a cognitive model of the thinking process.

Agreeing with the position of Urbina (2006) and Wilhelm (2005), the Inductive Reasoning Test for Children, or IRTC, Forms A and B – the object of the present study – was constructed based on the prescriptive theory of Inductive Reasoning Training by Josef Klauer (1990). Although this theory has been around for over 20 years, up to the present, no tests were found, specifically regarding inductive reasoning, which used this theory in the construction of items. Actually, this theory has been the basis for, mainly, intervention programs. In Brazil, Ávila (2002) adapted the (cognitive) psychopedagogical intervention program from Klauer to check the learning potential of children with and without learning difficulties. Before and after training with this cognitive program, the author applied the Raven test (that assesses intelligence, more specifically, inductive reasoning) and a dynamic intelligence test called LLT-BAK. The results showed an improvement in cognitive performance in all children evaluated by the LLT-BAK and Raven, in the application of these tests after training.

Thus, through his training theory, Klauer presented a clear conception about the cognitive processes involved in inductive reasoning and, from this, proposed training focused on learning how to reason inductively. The training program is the centerpiece of this theory, which is oriented to the thinking process, rather than to the product of thought. The research results are encouraging and can be found, principally, in Klauer (1990) and Klauer, Willmes, and Phye (2002); the former publication includes data from 30 studies. In all these, the training effectiveness was verified, with an increase of inductive reasoning, suggesting that it can be learned, because in the experiments, the group submitted to the intervention tended to present more than 1 standard deviation of increase in measures of reasoning, when compared to reference groups, which are the controls.

The inductive reasoning is a general result of individual observations, that is, it refers to the identification of regularities that are perceived in a given situation. For example, observing only white swans, inductively one could conclude that all swans are white, since the fact that all are white is a regularity. Therefore, the inductive reasoning starts from individual ideas that lead to general conclusions that may or may not be accurate. In the case of swans, the idea is incorrect, because although a human being might have only seen white swans, other colors also exist. However, in other situations, the inductive process may lead to correct conclusions (Klauer & Phye, 1995).

"The regularity represents an important role of thinking, because the regularities and uniformities provide the basis of the concepts and categories that serve as basic knowledge for abstract thought and reasoning" (Klauer & Phye, 1995, p. 37). The reasoning is inductive to detect regularities by the similarities and differences of attributes, or by relations with academic content, such as language, figures, drawings and numbers. This definition distinguishes the thoughts of inductive reasoning from the other types of reasoning, and a definition based upon a theory that specifies the cognitive processes that constitute inductive reasoning (Klauer & Phye, 1995).

Thus, Klauer (1990) classified the different types of inductive reasoning in relation to three facets: (a) if the reasoning is based on detection of similarities, differences, or both; (b) if the reasoning is based on detection of attributes of the stimuli or relations between them; and (c) if the stimulus, the object of reasoning, is of the linguistic kind, pictorial, geometric, numeric or otherwise. Based on the combination of the possibilities of the first two facets, Klauer and Phye (1995) categorized inductive reasoning into six related paradigms, namely: Generalization (similarity, attributes); Discrimination (differences, attributes); Cross Classification (similarity and differences, attributes); Recognizing Relationships (similarity, relations); Differentiating Relationships (differences, relations); and System Construction (similarities and differences; relations). Although this theory has been used primarily for development of intervention procedures, we chose to use it to support the creation of items of the IRTC, according to recommendations by Wilhelm (2005) regarding the importance of a solid theoretical framework to support the preparation of items for reasoning tests. Thus, the construction of items of Forms A and B of the IRTC were based on these paradigms and, for better understanding, each will be explained briefly below. Figurative examples of the paradigms are presented within *Methods*, by way of items developed for the IRTC.

Generalization (GE) refers to the need to establish similarities of attributes for different objects that form a group because, although there is differentiation, parts of these objects are common. To solve these problems the person proceeds analytically, comparing the properties of objects, trying to discover something in common between them.

Discrimination (DI) is the process of evaluating differences between objects in terms of attributes. Only one type of problem exemplifies this paradigm, i.e., to identify the object that doesn't belong to a group. This type of item also checks what is common among the objects, to therefore identify the difference. The contrast of this type of item with Generalization, previously presented, is that only one object here is unusual compared to the others. Again, to find the correct response, the attributes are compared to test the basic assumptions.

Regarding Cross Classification (CC), this type of task is organized in schemas of attributes in matrices of four parts (or four objects, two above and two below), where at least two attributes are considered simultaneously. For example, the object in the first quadrant must have at least one attribute equal to the object in the second quadrant, at its side, and another attribute equal to the third quadrant, which is below it. This should occur for each object in

each quadrant. The cross classification requires the determination of different and common attributes, and the advantage is that all the possibilities may occur, that is, similarity or difference in both the characteristics, or a mixture of similarity and differences.

In this type of item the subject also compares the objects, but crosses the information between all objects to find, for example, where a certain figure fits best. The item type used to this paradigm is presented with four objects in a matrix of four, one in each quadrant, and all presenting common attributes with two other objects. Another object is shown and the subject is asked in which quadrant it would fit best, leading the individual to analyze this outside figure and its attributes, comparing it with each figure in the quadrant until finding the quadrant that it would fit.

Recognizing Relationships (RR) is possible when at least two objects are present and inference of relation follows a procedure of comparison between pairs, but, in this case, it is necessary to find a similarity between relations rather than attributes, which makes resolving the item more complex. Recognition of relationships occurs in three types of problems: complementing the series, arranging the series, and analogies. For example, in analogy problems one must determine the specific relationship between a pair of objects that will be the benchmark. Several relationships may be possible, the solution strategy is to map the relationship of a pair of objects and establish, in an incomplete pair, this same type of relationship.

The paradigm Differentiating Relationships (DR) requires the recognition of differences in relations. This item type can be exemplified by only one type of problem, which are disturbances of the series, but can occur in two variations. In the first, the subject needs only to reorder the members of a problem set to correct the series, whereas in the second, an object should be excluded. In the two problem variations, the strategy is to find the relationship that occurs between the possible items, recognizing objects that distort the series.

Finally, the last paradigm is System Construction (SC). It involves a reasoning similar to that used for solving cross classification, but in the case of SC there are at least two relationships in which similarity and difference can be verified. Two types of problems represent this paradigm, a simple matrix, previously described, and the extended. The latter includes more than four objects in a set of problems and at least one pair of objects has a relation with another pair of objects, while there is relation differentiation with another pair of objects. To solve this type of item, one must recognize the operational relations and where there is similarity and difference.

Klauer and Phye (1995) also reported that the cognitive processes of Generalization and Discrimination paradigms occur in parallel, that is, to teach or apply one of these processes, the other tends to occur, because in the moment that one observes a resemblance, one must also notice

what is different to discover what is common between two or more objects. Thus, for the construction of the IRTC items, we chose to join the paradigms Generalization and Discrimination, as well as joining Recognizing Relationships and Differentiating Relationships. Therefore, rather constructing items within the six paradigms, four are referenced, being Generalization/Discrimination, Cross Classification, Recognizing/Differentiating Relationships, and System Construction. It is emphasized that the complexity of the items occurs along the paradigms, and gradually from the first to the sixth, Generalization/Discrimination, Cross Classification, Recognizing/Differentiating Relationships, and System Construction. Therefore, the least complex are less Generalization/Discrimination and the most difficult is System Construction.

From the paradigms described above, two parallel forms of the IRTC were created. The development of parallel forms is fundamental to the reapplication of the test, avoiding learning effects, and allowing us to assess the same skill, at two different times, with more reliability because the items in the Forms were constructed with the same parameters.

In this context, the present work was to conduct psychometric studies with both forms of the IRTC in order to investigate validity and reliability. The reliability was assessed using Cronbach's Alpha. With respect to evidence of validity (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 1999), internal structure and relationship with other external variables, specifically grade level, were explored. The first evidence allowed us to study the dimensionality of Forms A and B of the IRTC. The second, which refers to evidence-based relationship with other variables, had grade level as a criterion, because as it is a cognitive test addressing children's reasoning, it is assumed that there will be performance differences between the grade levels, given that beginners present lower performance than higher grade levels.

Method

Participants

The study participants included 417 children, 219 males and 198 females, sampled by convenience, being students from one public and one private school located in a city in the countryside of Sao Paulo state. Of the sample total, 167 were from public schools and 250 from private schools. In the public school sample, 87 subjects responded to Form A and 80 responded to Form B. In the private school sample, 132 responded to Form A and 117 to Form B.

Children attended 1st through 5th grade, with the application to 5th graders being only in the private school. Among the participant grade distribution, in public schools 23, 48, 42 and 54 children were assessed from

the 1st, 2nd, 3rd and 4th grades, respectively. In private schools, 58, 71, 49, 33 and 39 were assessed in the 1st, 2nd, 3rd, 4th and 5th grades. The mean age was 6.7 for 1st, 7.65 for 2nd, 8.86 for 3rd, 9.87 for 4th and 10.54 for 5th grade.

Instrument

Inductive Reasoning Test for Children-IRTC, Forms A and B (Muniz, Seabra, & Primi, 2008). The IRTC consists of two forms, A and B, each with 38 items, given that there are eight identical items in two ways. This composition was chosen because, with the use of Item Response Theory (IRT), one can calibrate the items in a single scale of difficulty. Thus, it is possible to obtain the parameters of the 68 items (30 + 30 + 8 anchors) on a common scale, avoiding all the subjects having to respond to all items, answering only 38 problems. The items in the two forms were designed to cover the four paradigms in accordance with Klauer's Theory and to develop balanced quantities of paradigms in each Form.

Each test item has four response options, but only one is correct. The items consist of pictorial and geometric figures. The correction utilizes 0 points for each incorrect answer and 1 point for a correct answer. The resulting scores are total and by paradigm. In the total, the correct responses are summed, out of the 38 items, and by paradigm, the correct answers are categorized and summed by Generalization/Discrimination (SD), Cross Classification (CC), Recognizing/Differentiating Relationships (RE), and System Construction (SC).

The maximum total in each test is 38, and by paradigm, the maximum on Form A is 8 in SD, 11 in RE, 9 in CC, and 10 in SC; and 9 in SD, 10 in RE, 9 in CC, and 10 in SC on Form B. In addition to the 38 items, each Form presents three items as examples, which are not counted as right or wrong. Figures 1, 2, 3 and 4 show an example of items prepared for each paradigm.
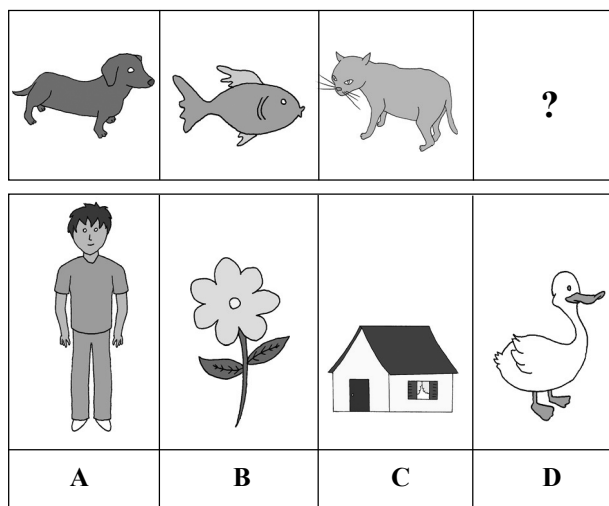

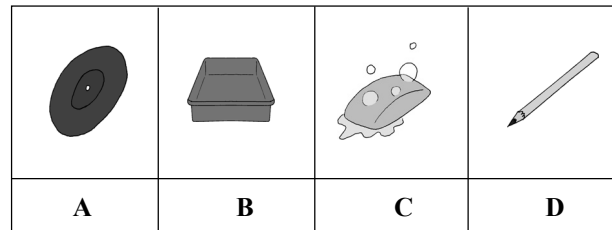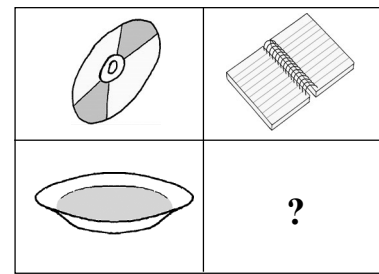*Figure 1*. Item of Generalization/Discrimination (SD) of attributes.

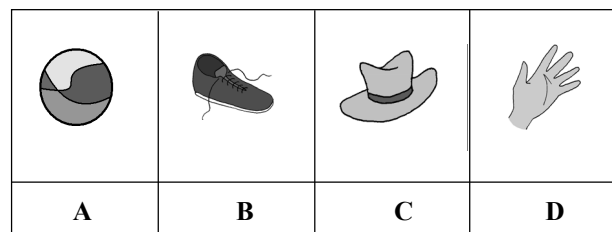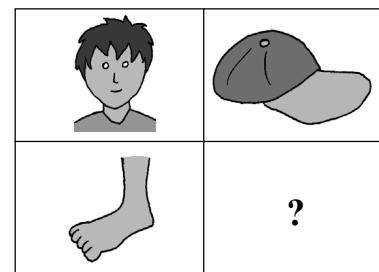*Figure 2*. Item of Cross Classification (CC).

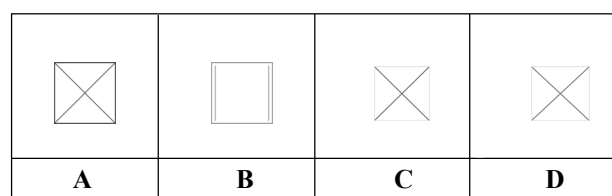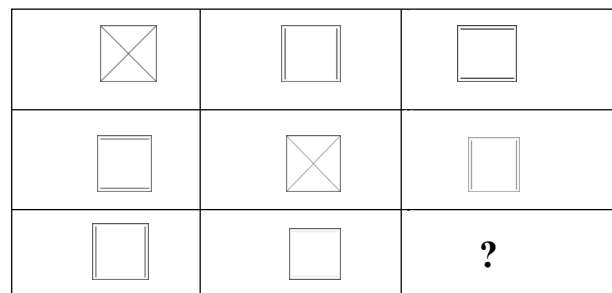*Figure 3*. Item of Recognizing/Differentiating Relationships (RE).

Figure 4. Item of System Construction (SC).

Procedure

First, permission was obtained from the Ethics in Research Committee, the consent of the schools and parents of the subjects; and consequently, Forms A and B of the test were applied to different samples, but equivalent in their characteristics such as grade and gender. The application was collective, in groups of approximately 30 students in each class, conducted by a trained psychologist and the first author of this article.

After the first contact, the test booklets were given to each child, with half the class responding to Form A and half to Form B, by random division. One of the proctors read the test instructions, which are the same for both Forms, to the children. Both Form A and B have the same three examples to be solved together with the children. After task understanding by the subjects, the test began and the answers were marked on an answer sheet. The proctors remained in the room to clarify any doubts and to check whether the children really understood the task. The application time ranged from 10 to 40 minutes, given that the lower grade levels were slower to respond.

Before starting the application, the teacher was asked about students that might potentially pose a problem that could interfere negatively on test responses, for example, in the case of mental or learning disorders. Within the sample, three dyslexic children were identified by information provided by the teachers. The answer sheets of these students were marked and later deleted from the database, because they could not solve the tasks, having been helped by the teachers.

Data Analysis

To study the reliability of Forms A and B of the IRTC, Cronbach's Alpha procedure was used, checking the associations among items within each form, based on the minimum reliability index of .60 to be considered satisfactory. In the evidence investigation of internal structure, seeking the dimensionality of Forms A and B the program Microfact was used, which makes use of IRT and tetrachoric correlation coefficients, using the scree plot as the structure to analyze the criterion found, the division between the eigenvalue of the first and second factor assuming values above 5 for suggested one-dimensionality and factor loadings of the items above .30.

Finally, the evidence from the relation with other variables, in this case, grade level, was verified by applying the ANOVA procedure of differences of mean, which indicates if there is significant difference in the data. However in order to identify among which grades the differences occurred, the Tukey analysis was applied.

Results and Discussion

Initially we investigated the internal structure of the scale. For this purpose we conducted a factor analysis, in order to verify that the structure of the IRTC Forms is similar to the quantity of six inductive reasoning paradigms proposed by Klauer or, in the case of the present study, with four paradigms, as explained in the introduction. This analysis aimed therefore to verify if Forms A and B of the test, compared to the items developed, would present uni- or multi-dimensional. For this we performed the factor analysis for complete information, with oblique rotation, through the Microfact program using the IRT and tetrachoric correlation coefficients, as they are suitable for items with dichotomous responses (Primi & Almeida, 1998). It is noteworthy that these analyzes were performed separately for Forms A and B.

Seeking to investigate a possible multi-dimensionality of the inductive reasoning construct, structured into four factors, according to the four paradigms, four factors were initially requested in the analysis for both Form A and Form B. We used the PROMAX rotation procedure, assuming an association between these factors. Then we examined the scree graphs verifying the plausibility of the extraction of the four factors. Also employed was the division of the eigenvalues of the first factor for the second, whose values higher than 5 indicate the one-dimensional structure (Hattie, 1985). In addition to these criteria, in each extraction, we examined the correspondence of groups of items with the model from the items with factor loading greater than .30 on each factor (Reise, Waller, & Comrey, 2000). Together these criteria suggested a correlated two-factor solution, and not four factors, for the two forms of the instrument.

It is noteworthy, through the literature retrieved, there is no data indicating whether or not paradigms are grouped into separate factors. Klauer (1990) proposes the theory, but the psychometric factor analysis to statistically test the theorization of the paradigms, it was not verified through the literature retrieved in the present study. Furthermore, the high correlation between different tasks measuring reasoning makes it difficult to divide into facets as defined. It is important to consider, however, that the absence of the separation factors does not invalidate the theoretical distinction with respect to the different types of processing involved in solving the problems presented. It should be noted that these paradigms are used to formulate the training activities of inductive reasoning, that is, in each paradigm the child learns to reason with more information and even using different procedures.

Analyzing the four factors obtained in both Form A and B, it was also not possible to verify a theoretical coherence among the items comprising each factor, but all correlated with each other in a positive form, and moderate or low. It is noteworthy that the first factor presents eigenvalue of 9.84 on Form A and 8.71 on Form B, while other factors presented values of approximately 2.5. Thus the ratio between the eigenvalues of the first and second factors are smaller than 5, indicating the unsuitability of a one-dimensional model. Table 1 shows the solution obtained after the PROMAX rotation of the two ways. For the

Table 1

Factors Extracted from Factorial Analysis Using the Program Microfact and Requesting Two Factors for Forms A and B

	Item Form A	Factor1	Factor2	Item Form B	Factor1	Factor2
1	A01SD	**.92**	-.12	C04SD_20	.28	.17
2	A02SD	**.90**	-.21	C06SD_02	**.60**	.20
3	A03SD_01	**.46**	-.04	C09CC	.12	-.17
4	C04SD_20	**.54**	.14	C12CC	.11	.15
5	A05SD	.04	.30	C19RE	.19	.37
6	C06SD_02	**.50**	.04	C22RE_27	**.87**	**-.33**
7	A07SD	-.01	**.37**	C25SC	**.78**	-.06
8	A08CC_05	-.01	**.54**	C28SC	**.39**	.08
9	C09CC	-.29	**.85**	B01SD_04	**.41**	.19
10	A10CC	.12	-.20	B02SD	.25	**.50**
11	A11CC_07	**.31**	.25	B03SD_19	**.50**	.16
12	C12CC	-.20	**.68**	B05SD	.12	.23
13	A13CC_24	**.38**	.26	B07SD_17	**.38**	.24
14	A14CC_23	-.00	**.58**	B08CC_22	.23	-.26
15	A15RE	**.60**	.09	B10CC	.04	-.03
16	A16RE	**.59**	.15	B11CC	.15	-.27
17	A17RE_10	**.70**	-.34	B13CC	.25	**.54**
18	A18RE	**.55**	.16	B14CC_08	**.35**	.07
19	C19RE	**.43**	.16	B15RE	**.66**	-.08
20	A20RE_11	**.64**	-.05	B16RE_09	.25	-.07
21	A21RE_12	.24	**.37**	B17RE_28	**.56**	.15
22	C22RE_27	**.48**	.24	B18RE	.06	**.34**
23	A23RE	**.46**	-.19	B20RE_25	.03	-.33
24	A24RE	**.50**	-.02	B21RE	.24	**.52**
25	C25SC	**.65**	-.05	B23RE	**.48**	.18
26	A26SC_16	**.75**	-.04	B24SC_32	**.58**	-.05
27	A27SC	-.10	.10	B26SC	**.59**	.08
28	C28SC	.29	.10	B27SC_14	**.40**	-.13
29	A29SC_29	**.37**	-.05	B29SC	**.65**	-.25
30	A30SC_31	.24	.24	B30SC_13	**.34**	0,01
31	A31SC	**.32**	.14	B31SC	**.68**	**.33**
32	A32SC_15	**.39**	.10	B32SC	**.29**	.00
33	A33SC	.18	.01	B33SC	.26	.12
34	A34SC_30	**.55**	-.07	B34SD_18	.12	**.54**
35	A35SD	**.61**	.11	B35SD_03	**.30**	.14
36	A36CC_21	**.41**	-.20	B36CC	-.03	.10
37	A37CC_06	**.75**	-.16	B37CC	-.10	**.48**
38	A38RE_26	-.02	.13	B38RE	.13	-.16
Eigenvalue		9.84	2.52		8.71	2.56
Explained Variance		24.34	4.80		19.71	4.60
Reliability		.84	.62		.85	.57

Note. The first letter of the Item allows for identification of which Form it belongs to: the letter A signifies Form A, the letter B signifies Form B, and the letter C signifies that the Item is common between both Forms. The number of items on the test is found after the paradigm type: in the case that the Item appears on both Forms, first is the number referring to its position on Form A and then a number referring to its position on Form B. The paradigm type is indicated by the alpha code, with code options for the four paradigms being: SD (Similarity/Difference), CC (Cross Classification), RE (Recognizing/Differentiating Relationship), and SC (System Construction).

earlier factorial analysis, structures were not described, because it was decided to describe only the structure that was most consistent.

Observing the structure of two factors in Form A, a more robust first factor was identified, with eigenvalue of 9.84, explained relative variance of 24.34 and con-

sisting of 25 items with factor loadings above .30, reaching reliability of .84. This factor is composed of representative items of all paradigms. Since factor two is less significant when compared with the first, with eigen-value of 2.52, explained relative variance of 4.80 and consisting of six items, obtaining reliability of .62. Note that this second factor is primarily composed of items of the CC paradigm. Despite checking this second factor, it was observed that the correlation between the two factors of .57 was considered moderately high. The explained relative variance is found by dividing the eigenvalues by the sum of the positive roots of the adjusted tetrachoric correlation matrix. Generally this sum is slightly different from the sum of items traditionally considered in the factorial analysis by principal component (Du Toit, 2003). Accuracies are calculated by the traditional formula of Cronbach's Alpha.

With respect to Form B, as described in Table 1, the two-factor analysis followed a pattern similar to that of Form A. We identified a more robust first factor, with eigenvalue of 8.71, explained variance of 19.71 and consisting of 19 items above .30, reaching an reliability of .85. The configuration of the structure also included items from all paradigms. Regarding the second factor, this was less expressive facing factor one, with eigenvalue of 2.56, explained variance of 4.60 and consisting of eight items that obtain reliability of .57. Of the items that composed the second factor, four were representative of the RE paradigm, two of SD and two of CC. There is no theoretical sense to justify this factor with such a structure. The correlation between the two factors was .38, which may be considered low, but almost moderate.

It was observed, however, that in both forms there was moderate to high correlation between the two factors and that the structures obtained from two factors are not theoretically expected. Therefore, we applied new factorial analysis, seeking only one factor. The results are shown in Table 2.

For Form A, in this new analysis, the eigenvalue and explained variance regarding the first factor remained the same as the previous analysis, but the factor has increased its membership to 30 items with loadings above .30 and the reliability improved to .87. With respect to the items, of the eight which did not achieve satisfactory load factor to be considered as belonging to the structure, three are part of factor two of the previous analysis. This data suggests that the form tends to present the one-dimensionality.

For Form B, as shown in Table 2, the eigenvalue and explained variance of factor one also practically did not change compared to the results of previous analysis, but the amount of items in factor one with loads above .30 increased to 25, and reliability remained stable at .87. This factor structure incorporated five of the eight items that were part of factor two in the previous analysis, and these items had factor loadings of moderate to high. As it was

for Form A, Form B was also best explained with the uni-dimensional structure.

From the data, Forms A and B are conceived as uni-dimensional since, despite having diverse items according to the paradigms proposed by Klauer (1990), the multi-dimensional structure comprised by factors corresponding to these paradigms was not observed. This seems to have been the first time the factor structure arising from the theory of Klauer has been investigated. The results found pointing to uni-dimensionality are nonetheless consistent with the theory mentioned, because all items are designed to be solved by inductive reasoning, and the main difference between the items is the complexity of the inductive thinking, but always from the point similarities between the information.

Josef Klauer (1990) proposes paradigms of inductive reasoning as if they are steps in the process of inductive thinking that are used depending on the situation that the subject faces. These paradigms, all based on the search for similarities between the information, have been created so that inductive reasoning was trained, particularly in children. Therefore, the paradigms were part of a single construct which was inductive reasoning, and were developed in the context of intervention, more than of evaluation. Although Klauer (1990) has differentiated the six paradigms for reasoning training, the author at no time stated that it was specific dimensions, distinct, which comprise inductive reasoning. The present study sought to test the operation proposed by Klauer, having adapted it to construct items of a psychological test that has been empirically tested. The results found in this present study help to better understand the theory of Klauer and contribute to the empirical and scientific advancement of the theory. The uni-dimensionality found in Forms A and B as well as the reliability obtained, suggest that the paradigm really assess inductive reasoning and that this construct, although it can be operationalized into facets, composes a single dimension.

It is noteworthy that, despite the factorial analysis having pointed to the uni-dimensionality of the Forms A and B, the descriptive statistics and correlation effected with scores separating each paradigm can be maintained, because the process of inductive reasoning has its particularity, as explained in the introduction. It is therefore interesting to see how the individuals behave in the face of sets of items in each paradigm. Furthermore, this study was treated as initial research, one should consider its limitations and the need for further exploration of results.

Despite the generally high accuracies obtained by PROMAX analysis, as the sample was composed of different grades, it is important to check the reliability separated by grade. Such information tends to clarify for which grade levels the Forms may be more reliable when applied. Table 3 may be referred to for these details, obtained through the analysis of Cronbach's Alpha.

Table 2
Factors Extracted from Factorial Analysis Using the Program Microfact and Requesting One Factor for Forms A and B

	Item Form A	Factor	Item Form B	Factor
1	A01SD	**.82**	C04SD_20	**.38**
2	A02SD	**.74**	C06SD_02	**.70**
3	A03SD_01	**.43**	C09CC	.01
4	C04SD_20	**.62**	C12CC	.20
5	A05SD	.24	C19RE	**.41**
6	C06SD_02	**.53**	C22RE_27	**.63**
7	A07SD	.24	C25SC	**.71**
8	A08CC_05	**.36**	C28SC	**.42**
9	C09CC	.21	B01SD_04	**.52**
10	A10CC	-.02	B02SD	**.55**
11	A11CC_07	**.48**	B03SD_19	**.59**
12	C12CC	.26	B05SD	.27
13	A13CC_24	**.56**	B07SD_17	**.52**
14	A14CC_23	**.39**	B08CC_22	.07
15	A15RE	**.66**	B10CC	.02
16	A16RE	**.69**	B11CC	-.01
17	A17RE_10	**.46**	B13CC	**.60**
18	A18RE	**.66**	B14CC_08	**.38**
19	C19RE	**.53**	B15RE	**.58**
20	A20RE_11	**.60**	B16RE_09	.20
21	A21RE_12	**.49**	B17RE_28	**.63**
22	C22RE_27	**.64**	B18RE	.27
23	A23RE	**.32**	B20RE_25	-.18
24	A24RE	**.49**	B21RE	**.57**
25	C25SC	**.61**	B23RE	**.58**
26	A26SC_16	**.72**	B24SC_32	**.53**
27	A27SC	-.02	B26SC	**.61**
28	C28SC	**.36**	B27SC_14	**.32**
29	A29SC_29	**.33**	B29SC	**.48**
30	A30SC_31	**.40**	B30SC_13	**.35**
31	A31SC	**.41**	B31SC	**.87**
32	A32SC_15	**.46**	B32SC	.29
33	A33SC	.18	B33SC	**.32**
34	A34SC_30	**.50**	B34SD_18	**.44**
35	A35SD	**.68**	B35SD_03	**.37**
36	A36CC_21	.26	B36CC	.02
37	A37CC_06	**.63**	B37CC	.19
38	A38RE_26	.07	B38RE	.02

The details found suggest satisfactory levels (above .60) and very satisfactory levels (above .80) depending on the grade – except for the 1st grade in both Forms, especially on B, for which there was no reliability, and the 2nd grade, also on Form B. Seeing that these reliability results, for the 1st grade on both Forms and the 2nd grade on Form B, are not satisfactory, the modification of some items is suggested or, after further study, the assumption that the test is not reliable for assessing children 1st and 2nd grades.

Finally, the validity evidence was checked, based on the relation with the grade level and the IRTC Forms. As the factorial analysis indicated uni-dimensionality, and because of the space restriction of the article composition, we chose to present this analysis of evidence, in more detail, only for the total scores of Forms A and B, although they could be done separately by paradigm. To investigate this type of validity, the mean values obtained in each grade were initially checked, and we were able to identify differences between the

Table 3

Reliability Index Evaluated for Form A and B by Grades

Form	Grade	Reliability
A	1ª	.48
A	2ª	.67
A	3ª	.80
A	4ª	.84
A	5ª	.75
B	1ª	No reliability
B	2ª	.58
B	3ª	.76
B	4ª	.77
B	5ª	.77

grades in both Forms, however, to see if the differences are statistically significant, we performed the ANOVA analysis of mean differences. The results showed significant differences among the grade level groups analyzed, namely, 1st, 2nd, 3rd, 4th and 5th grade scores of Forms A and B. To identify for which grades there were significant differences, Tukey analyses were conducted. The results can be seen in Tables 4 and 5.

Table 4

Differences in Mean Verified by Tukey Analysis between Each Evaluated Grade for the Total Score of Form A

Dependent variable	(I) Grade2	(J) Grade2	Mean difference	p
A_TT	1	2	-3.39 *	.04
		3	-5.80 *	.00
		4	-5.80 *	.00
		5	-11.34 *	.00
	2	1	3.39 *	.04
		3	-2.41	.25
		4	-2.41	.26
		5	-7.95 *	.00
	3	1	5.80 *	.00
		2	2.41	.25
		4	.00	1.00
		5	-5.54 *	.00
	4	1	5.80 *	.00
		2	2.41	.26
		3	.00	1.00
		5	-5.54 *	.00
	5	1	11.34 *	.00
		2	7.95 *	.00
		3	5.54 *	.00
		4	5.54 *	.00

*$p<.05$.

Observing Table 4 it appears that, for the total score of Form A, there is significant difference between grades, except between 2nd and 3rd, 2nd and 4th, and 3rd and 4th. This finding suggests that 2nd, 3rd, and 4th grades tended to obtain similar results in the total score of Form A. With this data it appears that Form A presents itself as sensible in differentiating 1st grade from the others, as is the case with 5th grade. These results, although not satisfactory as differentiation between the other grades was expected, are encouraging because the Form has already identified differentiated performance among children in less- and more-advances grades of elementary school.

The data obtained from Form B for different grades can be seen in Table 5.

Table 5

Differences in Mean Verified by Tukey Analysis between Each Evaluated Grade for the Total Score of Form B

Dependent variable	(I) Grade2	(J) Grade2	Mean difference	*p*
B_TT	1	2	-2.71	.13
		3	-7.20 *	.00
		4	-6.93 *	.00
		5	-9.70 *	.00
	2	1	2.71	.13
		3	-4.48 *	.00
		4	-4.22 *	.00
		5	-6.99 *	.00
	3	1	7.20 *	.00
		2	4.48 *	.00
		4	0.26	.99
		5	-2.50	.49
	4	1	6.93 *	.00
		2	4.22 *	.00
		3	-0.26	.99
		5	-2.77	.39
	5	1	9.70 *	.00
		2	6.99 *	.00
		3	2.50	.49
		4	2.77	.39

*$p<.05$.

As shown in Table 5, the ANOVA results with the grade level for Form B differ from the verified data for the Form A. On Form B, 5th grade did not differ significantly with grades 3 and 4. The same was observed between the 1st and 2nd grades. It was also noted that, on Form B, 3rd and 4th grade means were similar, but differed significantly from the 2nd grade. In this set of results, it is observed that Form B tended to differentiate between lower grades, the 1st and 2nd, and the more advanced, 3rd, 4th, and 5th. However, as on Form A, some items should be reviewed so that, in future studies, a greater differentiation between grades is found.

Final Thoughts

The purpose of this paper was to investigate some psychometric properties of Forms A and B of the Inductive Reasoning Test for Children-IRTC (Muniz et al., 2008). We specifically sought to identify the factorial structure of the Forms, the reliability and sensitivity for the sample differentiation of different school grade levels, in the case of 1st to 5th grades.

With regard to the factorial analysis, we could see the uni-dimensionality in the two Forms of the test, supporting the objective of Forms A and B, which is to measure the construct inductive reasoning. This result reinforces that the test tends to evaluate inductive reasoning in a purer fashion. This is an important fact which corroborates the theory that based the construction of each test item, given that they were developed to be reasoned inductively.

Although there was no differentiation in terms of the theoretically defined paradigms, it is important to highlight that Klauer's theory has contributed significantly to the development of different items, helping the broader mapping of the various processes that must be included in tests to assess inductive reasoning. This theory was chosen for the initial construction of the test and will continue to be for subsequent reformulations, by providing a consistent model to understand inductive thinking.

Another very positive outcome for Forms A and B were the accuracies found, the majority being satisfactory, above .60, with only the reliability for 1st grade on both forms, and 2nd grade on Form B being inadequate. Given this result, the information arising from the 1st grade and 2nd (this referring only to Form B), must be analyzed more carefully, seeing that the forms do not appear to measure the inductive reasoning of children in these grades as accurately.

Finally, the results obtained on the validity evidence with respect to other variables, in this case, grade level, provided some evidence of increased performance with the school progression on the two Forms. On Form A, grades 2, 3, and 4 showed similar scores, with grade 5 showing higher scores and grade 1 showing lower scores compared to the other three grades cited. With this, Form A seems to differentiate the 1st and 5th grade from the larger group composed of the 2nd, 3rd and 4th. Although there was no differentiation between the middle grades, on Form A, an improvement could be identified in the inductive thinking according to the increase of the grade

level. However, some items on Form A should be modified to try to differentiate between grades 2, 3, and 4.

Meanwhile, on Form B, the children in 1st and 2nd grade had lower scores in comparison to the 3rd, 4th, and 5th grades. Unlike Form A, on Form B the 1st and 2nd grades did not differ in scores, likewise with the 3rd, 4th, and 5th grades. Through Form B, a significant improvement in inductive thinking can be verified in the groups of 3rd, 4th, and 5th graders. But, as on Form A, the items should be improved to obtain greater sensitivity to assess differences between the majority of the grade levels.

The results found regarding the differences in scores between the grades legitimize the validity evidence by grade for Forms A and B, because they can identify significant differences in scores between grade level groups, and the higher grades obtained better results. This increase in scores by the advanced grades is expected when working with constructs related to intelligence, as is the case of inductive reasoning. Yet it is important that some items on the two Forms are reviewed in order to increase discrimination with the school progression.

Despite the psychometric information of Forms A and B being favorable, improvements are still needed relating to the reliability for the 1st grade, and 2nd grade for Form B, as well as score discrimination between the grades. Even with these improvements that must be shown, the IRTC already shows itself to be a differentiated test: the fact that it is a psychological test comprised of two Forms is important, as the practitioner will have more choices in testing, for example, using Form A for an initial evaluation and Form B for a re-evaluation. This procedure permits, in this situation, the application of two very similar tests, with the power to check, more reliably, for possible changes in the performance of the subjects, as well as the avoidance of the learning effect in skill testing when applying the same test twice. Thus, this article has contributed to the presentation of a test of inductive reasoning, developed based on a sound theory and with two options for testing the same ability. It is hoped that such initiatives can be fostered in the area of psychological assessment.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.

Ávila, M. M. (2002). *Avaliação do potencial de aprendizagem em crianças em processo de alfabetização*. (Dissertação de Mestrado não-publicada). Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, RS.

Du Toit, M. (Ed.). (2003). *IRT from SSI: BILOG-MG, MULTILOG, PARSCALE, and TESTFACT*. Lincolnwood, IL: Scientific Software International.

Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied and Psychological Measurement, 9*, 139-164.

Klauer, J. K. (1990). A process theory of inductive reasoning tested by the teaching of domain-specific thinking strategies. *European Journal of Psychology of Education, 5*(2), 191-206.

Klauer, J. K., & Phye, G. D. (1995). Cognitive training for children: A developmental program of inductive reasoning and problem solving. Toronto, Canada: Hogrefe & Huber.

Klauer, J. K., Willmes, K., & Phye, G. D. (2002). Inducing inductive reasoning: Does it transfer to fluid intelligence? *Contemporary Educational Psychology, 27*, 1-25.

Muniz, M., Seabra, A. G., & Primi, R. (2008). *Teste de raciocínio indutivo para crianças-TRIC-Formas A e B*. Itatiba, SP: Laboratório de Avaliação Psicológica e Educacional, Universidade São Francisco.

Primi, R., & Almeida, L. S. (1998). Considerações sobre a análise factorial de itens com resposta dicotómica. *Psicologia: Teoria, Investigação e Prática, 3*, 225-234.

Reise, S. P., Waller, N. G., & Comrey, A. L. (2000). Factor analysis and scale revision. *Psychological Assessment, 12*(3), 287-297.

Urbina, S. (2006). *Fundamentos da testagem psicológica*. Porto Alegre, RS: Artes Médicas.

Wilhelm, O. (2005). Measuring reasoning ability. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 373-407). London: Sage.