The construction and validation of a developmental test for stage identification: Two exploratory studies

Hudson F. Golino1, Cristiano Mauro Assis Gomes1, Michael Lamport Commons2, and Patrice Marie Miller3
1 Universidade Federal de Minas Gerais, Brazil
2 Harvard Medical School, USA
3 Salem State University, USA

The present work presents two exploratory studies about the construction and validation of the Inductive Reasoning Developmental Test (IRDT), a forty-eight items test based on the Model of Hierarchical Complexity. The first version of the test was administered to a convenience sample composed by 167 Brazilian people (50.3% men) aged between 6 to 58 years ($M = 18.90, SD = 9.70$). The Rasch Model was applied, and the result shows reliability of .97 for the full scale. The Infit mean was $87 (SD = .28; Max = 1.69; Min = .39)$, and the person reliability was .95. The one sample t-tests showed significant spacing of Rasch scores between items of adjacent orders of hierarchical complexity, with large effect size. The second study was conducted in order to overcome some of the test’s limitations found in the first study. The revised IRDT were administered to a convenience sample composed of 188 Brazilian people (57.7% women) aged between 6 and 65 years ($M = 21.45, SD = 14.31$). The reliability for the full scale was .99, and its Infit mean was $.94 (SD = .22; Max = 1.46; Min = .56). The person reliability was .95. The one sample t-tests showed significant spacing of Rasch scores between items of adjacent orders of hierarchical complexity, with large effect size. The paper finishes with a discussion about the necessity and importance to focus on the vertical complexity of the items in any test designed to identify developmental stages.

KEYWORDS: stages, assessment, validation, development, model of hierarchical complexity, inductive reasoning

Piaget is considered one of the most important researchers of the 20th century (Flavell, 1963), with his studies creating a very influential framework within developmental psychology, that of Genetic Epistemology. In spite of its importance, the influence of this theory on developmental research began to decline in the 1980’s, due to a large body of evidence that apparently contradicted the theory’s notion of developmental stages (Marshall, 2009; Miller, 2002). One might say that this theory was “put in check” by the maneuvers of others. When Piaget’s theory, specifically his stage concept, was put in check, all Piagetian and Neo-Piagetian developmentalists were, in some manner, placed in the same condition. As in chess, getting out of the check is of great importance, and requires the development and implementation of sturdy strategies. In developmental psychology, getting out of check can be reached through the implementation of “strategic moves”, as in the construction of better metrics (Fischer & Rose, 1999; Rose & Fischer, 1998; Van Geert & Steenbeek, 2005), with reliable, valid and accurate measures (Fischer & Dawson, 2002), and the adoption of quality control standards (Stein & Heikkinen, 2009).

The current paper presents one of these moves which, together with other works (Commons, Trudeau, Stein, Richards, & Krause, 1998; Commons et al., 2008; Dawson, 2003, 2006; Dawson & Wilson, 2004; Dawson, Goodheart, Wilson, & Commons, 2010; Dawson-Tunik, Commons, Wilson, & Fischer, 2005; Demetriou, Molenaar, De Boeck, & Van der Mass, 2005; Van der Maas & Kyriakides, 2006; Fischer, 2008; Fischer & Bidell, 1998, 2006; Rijmen, De Boeck, & Van der Mass, 2005; Van der Maas & Molenaar, 1992), aims to collaborate in getting out of the check. Two exploratory studies about the construction, challenges and initial results of the Inductive Reasoning Developmental Test (IRDT) - Teste de Desenvolvimento do Raciocínio Indutivo (Gomes & Golino, 2009) will be presented. The IRDT intends to measure the developmental stages of inductive reasoning through reliable, valid and accurate measures, falling in the category of so-called “quality control standards”.

Author note: Part of this research was supported by the Instituto Ester Assumpção, and by the Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG). We are thankful to all those involved in the revision of the manuscript. Special thanks to Prof. Cory David Barker (Antioch University Midwest), Prof. Igor Gomes Menezes (UFBA, Brazil) and Prof. Angela Maria Vieira Pinheiro (UFMG, Brazil), for all the suggestions and commentaries on the original manuscript.

Correspondence regarding this article should be addressed to Hudson F. Golino, Laboratory for Cognitive Architecture Mapping (LaiCo), Universidade Federal de Minas Gerais, Brasil. E-mail: hfgolino@gmail.com. Mobile: + 55 31 88607490
Criticisms of stages, or killing Piagetian stage theory:
Starting in the 1980’s, increasing numbers of researchers began to criticize Piagetian stage theory (Miller, 2002; Morra, Gobbo, Marin, & Sheese, 2008). The main criticisms were directed at the idea that stages are structures of the whole, developing in a synchronous way, emerging at specific ages, and reaching a single telos, represented by formal operations (Fischer & Bidell, 2006).


In addition to studies showing massive decalages, age issues and synchronism problems in Piagetian theory of cognitive development, other revisions of the theory were made. Commons and Richards (1984a), Commons, Richards and Kuhn (1982), Fischer (1980, 1987), Fischer, Hand and Russell (1984), and others, argued that the stage of formal operations is not the last possible level in human cognitive development, and show evidence for post-formal levels.

The other set of criticism emerged from philosophical/epistemological positions. Broughton (1984), for example, argued that formal operations are a wholly inadequate model of thought in adolescence and adulthood, and as a result suggests the entire theory should be reconsidered.

The criticism, sometimes based on empirical aspects, sometimes based on philosophical and epistemological positions, was striking, and came from many different lines. Flavell already in his early work entitled The Developmental Psychology of Jean Piaget (1961), pointed to ambiguities in the concept of stage, argued about the challenges of the clinical method, on the impossibility of stating that a child “has” a particular concept and raised the question of language as an intervening variable (Siegler & Crowley, 1991). Despite recognizing the historical importance of Piaget’s work, in particular the stage theory, Flavell comes to argue, in another, later work, that the Piagetian stage theory “explains nothing” (Flavell, 1985; Lourenço, 1998). Lourenço (1998) proposed that many cognitivists (e.g., Bjorklund, 1997; Brainerd, 1997; Cohen, 1983) already considered Piaget’s theory to be dead, and some of them suggested that there was no real purpose in continuing to test a theory that was already known to be inadequate (Halford, 1989; Lourenço, 1998).

In short, until the mid 80’s the classic structuralism of Piaget’s theory had significantly influenced developmental psychology research worldwide (Marshall, 2009). In spite of being one of the most important players of the “Developmental Chess,” the grandmaster was double checked. His influence, including the concept of stages, began to decline, due mainly to (1) the growing body of evidence that helped convince some researchers that stage theory was inappropriate to describe cognitive development (Morra, et al., 2008), and to (2) criticisms that addressed philosophical issues and suggested an epistemological reconfiguration (Marshall, 2009).

Neo-Piagetians and Post-Piagetians
A group of Neo-piagetian researchers has sought to overcome the problems and limitations pointed to in the Piagetian concept of stage, including his methodology for assessing them, proposing instead modern theoretical and methodological approaches that have been providing new evidences for discontinuity. Included in these new approaches are two important and related models of development: Fischer’s Dynamic Skill Theory (DST; Fischer, 1980; 2008) and Commons’ Model of Hierarchical Complexity (MHC; Commons, 2008; Commons & Richards, 1984; Commons & Pekker, 2008). Fischer (1980) proposed a set of analytical tools that make possible the detailed description of developmental pathways, as well as the construction of domain-free hierarchical taxonomies to classify performance. His DST (Fischer, 1980; 2008; Fischer & Bidell, 1998, 2006; Fischer & Rose, 1994, 1999; Fischer & Yan, 2002a, 2002b) conceives of development as a phenomenon composed of both continuous and discontinuous patterns of changes. The former (continuous change) relates to the sequence of steps followed in the construction of skills (microdevelopment) and the latter (discontinuous change) relates to abrupt, stage-like changes that marks the emergence of radically new kinds of control units of behavior and cognition (Fischer, 1980; Fischer & Rose, 1994; Fischer & Bidell, 1998, 2006; Fischer & Yan, 2002a). Evidence for both kinds of developmental patterns have been provided by Fischer and colleagues (Fischer, Kenny, & Pipp, 1990; Fischer & Silvern, 1985; Fischer & Yan, 2002a, 2002b; Schwartz & Fischer, 2005; Yan & Fischer, 2007). Instead of conceptualizing the discontinuous facet of human development as a unidirectional ladder the DST sees it as a constructive web that encompasses someone’s activity and the supportive context in which this activity is performed (Bidell & Fischer, 1992; Fischer & Bidell, 2006). So, a person may have a certain level of performance, let us say x, in the domain of Algebra, and an x-1 level of performance in the domain of Combinatorial Analysis, for example. Furthermore, this same person may present higher or lower levels of performance in the previously cited domains due to social support (scaffolding), emotional reactions, and so on (Fischer & Bidell, 2006). The constructive web notion is different from the Piagettiad concept of stages as developmental ladder, in which decalage is the exception.

Despite the importance and contribution of the DST to the Developmental Sciences field (Miller, 2002; Morra et al., 2008), it was Commons and his colleagues that have proposed the groundwork for the mathematical formalization of discontinuity, through the Model of Hierarchical Complexity (MHC). The MHC is a general measurement theory, and as such is part of the normal Mathematical Theory of Measurement (Krantz, Luce, Suppes, & Tversky, 1971; Luce, & Tukey, 1964) applied to the phenomenon of difficulty. The MHC introduces the concept of the Order of Hierarchical Complexity (OHC) that conceptualizes information in terms of “the power required to complete a task or solve a
problem” (Commons, Trudeau, Stein, Richards, & Krause, 1998). Commons and Pecker (2008) demonstrated, in axiomatic terms, that task difficulty or complexity, beyond other sources, increases in two ways: horizontally and vertically. The first refers to the accumulation of informational bits necessary to successfully complete a task (Commons, 2008), e.g. \(5 + 6 + 7\) is less complex than \(5 + 6 + 7 + 8\), because the first differs from the second in the number of times addition was executed, and does not differ in the organization of the addition itself; that is, both have the same hierarchical (or vertical) complexity. So, horizontal or traditional complexity is just the adding of informational bits. Vertical complexity, or hierarchical complexity, refers to the organization of information in the form of action in two or more subtasks, in a coordinated way. The distributive property is a good example of vertical complexity. Let’s take the following example: \(a \times (b + c) = (a \times b) + (a \times c)\). In order to correctly perform the task, one should multiply the element \(a\) by \(b\) and by \(c\), separately, and then sum the results, or sum \(b\) with \(c\), and then multiply by \(a\). If someone change the order of execution of the actions, e.g. \((a \times b) + c\), the result won’t be right. So, it requires the two actions of addition and multiplication to be performed in a certain order, thus, coordinated.

Briefly summarizing, the MHC postulates that actions at a higher order of hierarchical complexity: 1) are defined in terms of two, or more, lower-order actions; 2) organize and transform those actions, not just combine them in a chain; and 3) produce organizations of lower-order actions that are new and not arbitrary. The first two are also Piagetian postulates, but the third is not. The order of hierarchical (or vertical) complexity refers to the number of recursions that the coordinating actions must perform on a set of primary elements (Commons, 2008). Because hierarchical complexity is a property of tasks, performance is separated from tasks. Stage is defined as the most hierarchically complex task solved. Each task that occurs in a separate domain is considered separately. There is no structure of the whole, so in the DST, decalage is the normal modal state of affairs.

The development of metrics in developmental psychology has been one of the challenges and needs of the area (Van Geert & Steenbeek, 2005; Fischer & Rose, 1999), and it is considered crucial in guiding research and professional practice (Stein & Heikkinen, 2009). The Hierarchical Complexity Score System – HCSS (Commons, LoCicero, Ross & Miller, 2010; Dawson, Commons, Wilson, & Fischer, 2005) and the Lectical Assessment System – LAS (Dawson-Tunik, 2004) represent general, reliable, valid, domain-free scales or metrics (Dawson, 2004). These metrics were studied by Dawson (2000, 2001, 2002, 2003, 2004) who compared them with domain-specific scales, such as the Good Life Scoring System (Armon, 1984), the Standard Issue Scoring System (Colby & Kohlberg, 1987a, 1987b) and the Perry Scoring System (Perry, 1970). Dawson (2005) points out that, in spite of measuring the same latent variable, the domain-free scales present better internal consistency, allow meaningful comparisons across domains and contexts, and enable the examination of the relationship between developmental stages and conceptual content. Moreover, the HCSS and the LAS are considered two of few calibrated developmental metrics in use, being studied in terms of their construct and congruent validity, internal consistency and inter-rater reliability, providing evidence of fine grained interval scales (Stein & Heikkinen, 2009).

Despite the importance in guiding developmental and psycho-educational research and practice, the domain-specific scales demand various trained scoring analysts, with high agreement between them, require a considerable time for large scale evaluation and are vulnerable to subjective bias. So, the construction of objective large-scale tests can help the field to move beyond these challenges, bringing speed and lower-cost procedures for evaluating discontinuities.

### Table 1. Some instruments based on the model of hierarchical complexity and/or dynamic skill theory

<table>
<thead>
<tr>
<th>Problem-solving</th>
<th>Vignettes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algebra (Richardson &amp; Commons, 2008)</td>
<td>Social perspective-taking (Commons &amp; Rodriguez, 1990; 1993)</td>
</tr>
<tr>
<td>Infinity (mathematics) (Richardson &amp; Commons, 2008)</td>
<td>Attachment and loss (Miller &amp; Lee, 2000)</td>
</tr>
<tr>
<td>The laundry problems (Goodheart &amp; Dawson, 1996; Goodheart, Dawson, Draney, &amp; Commons, 1997)</td>
<td>Workplace organization (Bowman, 1996a; 1996b)</td>
</tr>
<tr>
<td>The combustion problem (Bernholt, Parchmann, &amp; Commons, 2008)</td>
<td>Workplace culture (Commons, Krause, Fayer, &amp; Meaney, 1993)</td>
</tr>
<tr>
<td>The moral decision problem (Commons, Richards &amp; Kuhn, 1982)</td>
<td>Political development (Sonnert &amp; Commons, 1994)</td>
</tr>
<tr>
<td>The Index of Mortality Problem (Dawson, 2004)</td>
<td>Relationships (Armon, 1984a)</td>
</tr>
<tr>
<td>The decision problem (Commons, Lee, Gutheil, Goldman, Rubin, Appelbaum, 1995)</td>
<td>Views of the “good life” (Danaher, 1993; Dawson, 2000; Lam, 1994)</td>
</tr>
<tr>
<td>Loevinger’s sentence completion task (Cook-Greuter, 1990)</td>
<td>The student-bully problem (Joaquim, 2011)</td>
</tr>
<tr>
<td>Report patient’s prior crimes (Commons, Lee, Gutheil, Goldman, Rubin, Appelbaum, 1995)</td>
<td>Causing religious beliefs / causing atheism (Miller, Harrigan, Commons, &amp; Commons-Miller, 2008)</td>
</tr>
<tr>
<td>Causing religious beliefs / causing atheism (Miller, Harrigan, Commons, &amp; Commons-Miller, 2008)</td>
<td>Other</td>
</tr>
</tbody>
</table>

**Other**

- Four story problem (Commons, Richards & Kuhn, 1982; Kaillo & Heikama, 1991)
- Counselor stages (Lovell, 2002)
- Loevinger’s sentence completion task (Cook-Greuter, 1990)
- Report patient’s prior crimes (Commons, Lee, Gutheil, Goldman, Rubin, Appelbaum, 1995)
- Causing religious beliefs / causing atheism (Miller, Harrigan, Commons, & Commons-Miller, 2008)
- The student-bully problem (Joaquim, 2011)
The MHC can be used not only to construct analytic scales, but also for the construction and design of tests, tasks and vignettes. A number of tasks have been created in many domains, based on the MHC or DST (as seen in Table 1).

Constructing calibrated tests for developmental stage identification requires a specific design as defined by Commons and colleagues (Commons & Pekker, 2008; Commons, Gane-McCalla, Barker & Li, this issue). This design involves: 1) grouping items with same hierarchical complexity \[ h(i_1) = h(i_2) = \cdots = h(i_n) \] within stages; and 2) using items with increasing hierarchical complexity \[ h(\varphi_1) < h(\varphi_2) < \cdots < h(\varphi_k) \] within stages. The first deals with item or task equivalence, important in order to avoid the elaboration of an anomalous scale that confuses its analysis (Fischer & Rose, 1999). The second makes possible the identification of discontinuous, stage-like development, with gaps between different orders. There is an expected item structure of any instrument construct based on the MHC. That structure focuses on both strategies in order to identify developmental stages should be as close as possible to the diagram below (Fig. 1). Each blue box in the Figure 1 represents a cluster of items of the same hierarchical complexity or OHC. Within a single box, the items have the same Order of Hierarchical Complexity (h) in that domain. The OHC of the items increases from stage 1 (\( \varphi_1 \)) to stage \( k(\varphi_k) \), so that \( h(\varphi_1) < h(\varphi_2) < \cdots < h(\varphi_k) \) (Consequences 2, 3 and 4 of the formal MHC; see Commons & Pekker, 2008; Commons, Gane-McCalla, Barker & Li, this issue). Furthermore, the figure shows the expected gaps between the clusters of adjacent OHC items (see Figure 1).

Beyond the strategies of grouping items with same OHC and using items with increasing OHC, in order to identify developmental stages, a good measure or ruler needs to address a single trait or dimension, be constructed based upon an explicit theory or model of development (Stein, Dawson & Fischer, in press), be submitted to empirical investigation, aiming to test the expected equivalence and order of items, and determine other scale properties (Fischer & Dawson, 2002; Fischer & Rose, 1999). Commons and colleagues (Commons et al., 2008; Dawson, Goodheart, Draney, Wilson, & Commons, 2010) evaluated the expected equivalence and order of items from the developmental test design through the Rasch family of models (Andrich, 1988; Rasch, 1960). The dichotomous Rasch Model (Rasch, 1960/1980), also called Simple Logistic Model (SLM) for dichotomous responses (Andrich, 1988), establishes that the right/wrong scored response \( X_{vi} \), that emerges from the encounter between the person \( v \) and the item \( i \), depending upon the performance \( \beta \) of that person and on the difficulty \( \delta \) of the item. Its relation can be expressed as the following probabilistic function:

\[
P\{ X_{vi} = x \} = \frac{e^{x(\beta_i - \delta_i)}}{1 + e^{(\beta_i - \delta_i)}}
\]

The Rasch model deals with the relationship between the person ability and item difficulty in a probabilistic way. Both parameters are allocated on a single abstract continuum that goes from "low" to "high" ("more" or "less", etc), concerning just one attribute of the object (or attitude, or behavior) measured, thus unidimensional. In the Classical Test Theory (CTT) the corresponding "parameter" for the Rasch's person performance (\( \hat{\beta} \)) is the estimated true score (\( T_v \)), or the score reported on test-score scale (normally distributed) (Hambleton & Jones, 1993). It can indicate the "position" of the person on the construct measured, but unlike the SLM, needs a representative sample for unbiased item estimates, a norm group.

---

**Figure 1.** Expected Item structure of instruments constructed focusing on the vertical complexity within a specific domain (unidimensional).
for comparison between individuals, giving meaning to the scores, and a normally distributed score for achieving interval scales properties (Embretson & Reise, 2000).

Some authors argue that the dichotomous Rasch model is the simplest Item Response Theory model (one-parameter model) (Bock & Jones, 1968; Hambleton, 2000). However, Andrich (2004) argues that differently from the traditional IRT paradigm, in which one chooses the model to be used (one, two or three parameters) according to which better accounts for the data, in the Rasch Paradigm “the SL.M is used because it arises from a mathematical formalization of invariance which also turns out to be an operational criterion for fundamental measurement” (p.15). So, instead of data modeling, the Rasch’s paradigm focuses on the verification of data fit to a fundamental measurement criterion, compatible with those found in the physical sciences (Andrich, 2004, p.15).

From among the benefits of using the Rasch family of models for measurement, some should be highlighted. In sum, it allows the construction of objective and additive scales, with equal-interval properties (Bond & Fox, 2001; Embretson & Reise, 2000), it produces linear measures, gives estimates of precision, allows the detection of lack of fit or misfit and enables the parameters’ separation of the object being measured and of the measurement instrument (Panayides, Robinson & Tymms, 2010). It also makes possible the reduction of all of a test’s items into a common developmental scale (Demetriou & Kyriakides, 2006), collapsing in the same latent dimension person’s abilities and item’s difficulty (Bond & Fox, 2001; Embretson & Reise, 2000; Glas, 2007), and enables the verification of hierarchical sequences of both item and person, being especially relevant to developmental stage identification (Dawson, Xie & Wilson, 2003).

Through the assumptions and procedures introduced by Commons and colleagues (Commons and Pekker, 2008; Commons et al., 2008; Dawson–Tunik et al., 2010) it has become possible to design and construct valid and reliable developmental metrics, tests and tasks, bringing new empirical evidence that helps reveal stage-like discontinuity. Following this tradition, two exploratory studies about the construction, challenges and initial results from the construction of an objective, large-scale instrument, named the Inductive Reasoning Developmental Test (IRDT), developed by Gomes and Golino (2009). These studies will be presented in some detail with the aim of unpacking the challenges involved in the construction of a developmental test, and will present a methodology for developmental stage identification. This methodology is put forward as one of the moves that can help uncheck the idea of stages within the virtual game of “Developmental Chess”, together with other moves published elsewhere (Demetriou & Kyriakides, 2006; Rijmen, De Boeck, & Van der Mass, 2005).

Study I: Uncovering discontinuities, and finding alternative sources of difficulty beyond vertical complexity

The purpose of Study 1 was to construct the initial version of the instrument, and in so doing, assess the scale structure of the items, verifying if they presented previously predicted orders and gaps, and to investigate the initial estimates of reliability and unidimensionality, among other scale properties, using Rasch analysis.

### Figure 2.

<table>
<thead>
<tr>
<th>Item</th>
<th>Letters</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
</tr>
</tbody>
</table>

The IRDT (Gomes & Golino, 2009) is a pencil-and-paper instrument designed to assess developmentally sequenced and hierarchically organized inductive reasoning. It is an extension, in terms of complexity, from the Indução test, which compose the fluid intelligence test kit (Gomes & Borges, 2009) of the Higher-Order Cognitive Factors Kit (Gomes, 2010). The domain of inductive reasoning was used because it is one of the best indicators of fluid intelligence (Carroll, 1993). The construction of the IRDT, from the original Indução items, is due to a larger challenge that concerns the construction of an intelligence battery to identify developmental stages.

The sequence of IRDT was constructed based on the MHC and on Fischer’s Dynamic Skill Theory. It was designed to identify six developmental stages (or levels), that will be named based in both theories, respectively: Pre-operational or Single Representations (Pre-op/sr); Primary or Representational Mappings (Prim/rM); Concrete or Representational Systems (Conc/rs); Abstract or Single Abstractions (Abst/sA); Formal or Abstract Mappings (Form/AM); and Systematic or Abstract Systems (Syst/sA). Each stage is composed of eight items with the same order of hierarchical complexity (OHC), for a total of forty-eight items. Each item is composed of four letters, or sequence of letters, with a specific rule (correct items), plus one letter or sequence with a different rule (exception). The task is to discover which letter or sequence is the exception. From stage to stage, there is a difference of +1 in the Order of Hierarchical Complexity (OHC). The instructions for performing the test is as follow: “You’ll be presented several reasoning tasks (items). In each task (item) you have five letters or sequence of letters. Among the five letters or sequence of letters, four of them have a specific rule, and one has a rule that is different from the others. Your challenge is to identify (marking with an X) the letter or the sequence of letters that has a different rule, compared to the other four. Each task (item) is displayed in a specific row, beginning with a number, from 1 to 48. You have no time limit. Solve as many tasks (items) as you can.”

**Pre-operational or Single Representations (Pre-op/sr):** Each item is composed of specific letters. The rule is “equal letter”, and the exception is a different one (see Figure 2).

**Primary or Representational Mappings (Prim/rM):** Eight items were created for this stage. Four of them have a specific rule: there is no jump in the letters’ sequence. In the example below, the first option is composed of wx. There is no other letter between them, so they form a non-jump sequence (Rule 1). The exception, however, is a conjoint of two letters that jumps one letter of the alphabetic sequence (e.g. qs; see Figure 3).

### Figure 3.

| Example: Item Prim/MR1 – Rule 1 |
|------|------|------|------|------|------|
| 9    | WX   | MN   | ST   | QS   | YZ   |
METHOD

Participants

In Study 1, the IRDT was administered to a convenience sample composed by 167 Brazilians (50.3% men, 49.7% women) aged between 6 to 58 years (M = 18.90, SD = 9.70). The sample was intentionally broad, and had a distribution of 15.6% from 6 to 12 years, 27.5% from 13 to 15 years, 35.9% from 16 to 20 years, and 21% beyond 20 years. All the participants were from the city of Belo Horizonte, state of Minas Gerais, Brazil.

Procedure

The data were collect by the first author and by thirty Psychology undergraduate students, enrolled in a first semester Cognitive Development class, trained in how to administer the instrument properly. The author first administered the instrument to the undergraduate students (whose data are being used in this analysis), and to 47 first year high school students from a public school. Each undergraduate student was assigned to administer the IRDT to three different people from 6 to 60 years of age. Participation was voluntary, with people agreeing to be part of the study after its purpose was explained. They were informed that their answers would be kept confidential, and that all procedures guaranteeing the privacy of their results would be adopted. They then signed an inform consent form, as required by the guidelines of the Ethical Committee of the Universidade Federal de Minas Gerais, Brazil.

Data analysis

In the first part of the data analysis the dichotomous Rasch Model is used, making it possible to reduce the items from the IRDT into a developmental scale (Demetriou & Kyriakides, 2006), collapsing at the same level person’s abilities and item’s difficulty (Bond & Fox, 2001; Embreeston & Reise, 2000; Glas, 2007). It also enables...
the verification of hierarchical sequences of both item and person, being especially relevant to developmental stage identification (Dawson, Xie & Wilson, 2003).

To verify the adjustment of the data to the model, the Infit (information-weighted fit) mean-square statistic is used. It represents "the amount of distortion of the measurement system" (Linacre, 2002, p.1). Values between 0.5 and 1.5 logits are considered productive for measurement, and <0.5 and between 1.5 and 2.0 are not productive for measurement, but do not degrade it (Wright & Linacre, 1994). The unidimensionality of the instrument can be checked by a number of procedures, each one complementing the other (see Tennant & Pallant, 2006). Here, unidimensionality will be addressed using only the model fit statistics – i.e. if the data fit the model, one of the consequences is the linearity of the measure, its unidimensionality, and so on – and the principal contrast, which can be verified through the percentage of variance explained by measures, and by the percentage of unexplained variance in the first contrast. The former should be closer to or greater than 60% (Peeters & Stone, 2009), while the latter should be closer to or less than 10%.

In the second part of the analysis, the spacing of Rasch scores between items of adjacent orders of hierarchical complexity is described. The Rasch scores represent the difficulty of an item (δ), which is its location at the latent variable continuum. It would have been good to compare the Rasch Scores for every item from adjacent orders of hierarchical complexity, but because there were so many items, this would have produced too many comparisons. To reduce the number of comparison pairs, each item's Rasch score was subtracted from the mean Rasch score of the items from the next higher order of complexity. This calculation is represented by the Formula 2:

$$\overline{X}_{k+1} - \delta_i = Adj \delta_{ik}$$ (2)

where $\overline{X}_{k+1}$ is the mean of the next higher order of complexity (or Stage $k+1$), and $\delta_i$ is the difficulty of item $i$ from order $k$ (or Stage $k$), producing the adjusted difficulty of item $i$. To verify if the differences between difficulties of items from order $k$ and the mean difficulty of the order $k+1$ are statistically significant, the One-Sample t-test is used, with a 95% confidence interval. The effect size ($r$) is calculated using the Cohen's $d$ and effect size correlation $r$ (Rosnow & Rosenthal, 1996).

**RESULTS**

The Rasch dichotomous model (Andrich, 1988; Rasch, 1960) was calculated using the software Winsteps (Linacre, 1999, 2011). Out of the 48 items, 5 were responded correctly by all participants (Pre-op/SR, Pre-op/SR3, Pre-op/SR4, Pre-op/SR5 and Pre-op/SR8). The reliability for the forty-three non-extreme items was .99, and for the full scale (48 items) the reliability was .97. The Infit mean was .87 ($SD = .28; Max = 1.69; Min = .39$), falling within the acceptable fit range. The person reliability was .95, which is estimated to indicate the degree to which a person's response pattern conforms to the difficulty structure of the measure (Hibbard, Collins, Mahoney & Baker, 2009). The principal contrast showed that the raw variance explained by measures (modeled) is 70.6%, and that the unexplained variance in the first contrast (modeled) is 10.4%, suggesting that the instrument can be thought of as unidimensional.

The variable map (Figure 14) illustrates the scale for the 48 items of the IRDT with item difficulties (on the right) and person measures (on the left) calibrated on the same scale. It is visually possible to identify clear item clusters in the Systematic/Abstract Systems’ stage (Syst/AS1, Syst/AS2, Syst/AS3, …, Syst/AS8) and in the Formal/Abstract Mappings’ stage (Form/AM1, Form/AM2, Form/AM3, …, Form/AM8), with a gap between them. The Abstract/ Single Abstraction’s items presented a cluster (they are all together without any other stage’s items), but did not present a gap in relation to the Concrete/Representational System’s items. Some Primary/Representational Mappings’ items (Prim/rm5, Prim/ rm6, Prim/rm7, Prim/rm8), had difficulties very close to the Concrete/rs’s items, making one big item set. The other Primary/ rm’s items (i.e. Prim/rm1, Prim/rm2, Prim/rm3 and Prim/rm4) were less difficult than other items of the same stage. Moreover, they presented a gap in relation to the item’s set composed by the other Primary items and by the Concrete ones. Finally, the relative position of person (left) and item (right), shows the IRDT as an easy test for 23 participants ($Mean\ ability = 7.66, SD = 0.81$). The whole sample mean ability was 1.15 with standard deviation of 3.40 logits (see Figure 14).

**Table 2.** One-sample tests of mean item difficulties for different OHC’s

<table>
<thead>
<tr>
<th>Stages</th>
<th>Test value = 0</th>
<th>95% CI</th>
<th>Effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$t$</td>
<td>$DF$</td>
<td>Sig. (2-tailed)</td>
</tr>
<tr>
<td>Pre-op/SR and Primary/RM</td>
<td>13.58</td>
<td>7</td>
<td>0.00</td>
</tr>
<tr>
<td>Primary/RM and Concrete/RS</td>
<td>3.29</td>
<td>7</td>
<td>0.01</td>
</tr>
<tr>
<td>Concrete/RS and Abstract/SA</td>
<td>7.99</td>
<td>7</td>
<td>0.00</td>
</tr>
<tr>
<td>Abstract/AS and Formal/AM</td>
<td>36.01</td>
<td>7</td>
<td>0.00</td>
</tr>
<tr>
<td>Formal/AM and Systematic/AS.</td>
<td>9.49</td>
<td>7</td>
<td>0.00</td>
</tr>
</tbody>
</table>

**Figure 8.** Example: Table Row 1, Abstract/SA

**Figure 9.** Example: Table Row 2, Abstract/SA
DISCUSSION

The current study aimed to assess the scale structure of the items, verifying whether they represented previously predicted orders and gaps (see Fig.1), and to investigate the initial estimates of reliability and unidimensionality, among other scales properties, using Rasch analysis. The result suggests the unidimensionality of the items, to some extent, since the percentage of raw variance explained by the measures (modeled) is moderately high (70.6%), and the principal components analysis of the residuals gave an unexplained variance of 10.4% for the first contrast. The items’ adjustment to the model was verified through the Infit index, which was found to have a mean of .87 and a standard deviation of .28. The minimal Infits value was .39 (Item System/AS.4) and the maximum was 1.69 (Item Primary/MR.5), and all other non-extreme items had Infits smaller than 1.32. This is considered to reflect a good fit to the model. The person and item reliabilities were good (.97 and .95, respectively). After assessing some of the psychometric properties of the measures, it was necessary to look more closely at the variable map (Fig.1).

The Pre-operational/Single Representation stage presented two sets of item difficulties, i.e. items Pre-op/sr1, Pre-op/sr3, Pre-op/sr4, Pre-op/sr5 and Pre-op/sr8 were shown to be less difficult than items Pre-op/sr2, Pre-op/sr6 and Pre-op/sr7. This gap between items with the same predicted ÖHC suggests that there was a problem in designing these items. One hypothesis to explain this effect could be that they are more horizontally complex. The Preo-operational items are composed of four equal letters plus a different letter, requiring the participant only to discriminate a set of five simple stimuli, choosing the dissimilar one. The items Pre-op/sr2, Pre-op/sr6 and Pre-op/sr7 may have been more difficult because the letters provided as options, in each item, were closer in graphical terms. The item Pre-op/sr2, for example, was composed by four "O" and one "Q". The visual stimuli of both letters are graphically closer, differing by the little “dash” on the bottom of Q. Previous research has shown that the structure of cognitive processing is composed of cascade-like relations (Demetriou, Christou, Spanoudis, & Platsidou, 2002; Demetriou, Mouyi, & Spanoudis, 2008) between processes with increasing complexity, beginning with speed processing (the most basic component of the cognitive architecture), followed by perceptual discrimination, perceptual control, conceptual control, short-term memory, working memory and, finally, reasoning processes. According to Demetriou, Mouyi and Spanoudis (2008), perceptual discrimination “reflects sheer speed of processing together with the processes required to discriminate between two simple stimuli and identify the target one” (p. 439). So, when comparing different stimuli, those whose difference are based on small tiny cues (e.g. the little dash of letter Q), demand a higher perceptual discrimination than those having more cues (e.g. comparing “A” with “E”). Thus, Pre-op/sr2, Pre-op/sr6 and Pre-op/sr7 are more horizontally complex than the other four Pre-operational items, because they demand a slight higher level of perceptual discrimination. In sum, it seems that in items from the Pre-operational order it is important to control as much as possible the perceptual discrimination required for the item or task, in order to avoid interference from the standpoint of horizontal complexity.

The next order’s items also present two set of difficulties. The items Prim/rim1, Prim/rm2, Prim/rm3 and Prim/rm4 were the easiest items of the Primary stage, probably because they were constructed according to the Rule 1, i.e. four options with no jump between the pair of letters, and one option jumping one letter. The other four Primary items where constructed according to the Rule 2, which states a jump of one letter between each pair of letters (4 options), and one option jumping two letters. Our hypothesis is that when dealing with items constructed according to Rule 2, the participants needed to store and deal with more information in Working Memory (Demetriou et al., 2002, 2008; Pascaul-Leone, 1984), which could horizontally increase the complexity of the task. A similar effect also seems to occur with the next order’s items. Note the items Conc/r5, Conc/rs6, Conc/rs7 and Conc/rs8, which are the most difficult concrete items, have a mean difference of .92 logits from the Conc/rs1, Conc/rs2, Conc/rs3 and Conc/rs4. This might be because the most difficult items have a rule which involves one more bit of information, being more horizontally complex than the items Conc/rs1, Conc/rs2, Conc/rs3 and Conc/rs4. Originally, we varied some of the rules somewhat in order to make the task less boring, and to avoid possible fatigue from the repetition of procedures employed to answer an item or task. However, our result suggests that changing some items’ rules within a certain ÖHC can compromise the quality of the stage identification. It seems that a good strategy for developmental test construction is trying always to elaborate items with the same rule within a single ÖHC.

The items from the Abstract, Formal and Systematic orders, on the other hand, are forming groups, or clusters, reflecting the fact that items within each are of the same hierarchical complexity (and are therefore grouped together), and items across each order are appropriately separated. The abstract items, however, are not well separated from the Concrettes items. It can be speculated that the way the tables of the Abstract order were constructed, having eight code rows, each beginning with an alphabetic letter followed by a Greek letter, decreases the difficulty of the items. The options of the items are all organized and well structured, and this organization seems to work as a support for the respondents.
In spite of providing good indicators of the items’ structure, and enabling the verification of visual clusters of items, the Rasch analysis did not provide information regarding the size of the gaps between adjacent OHC’s. The one-sample t-tests, calculated for this purpose, showed that the differences between adjusted difficulties of items from adjacent orders are statistically significant, with large effect sizes. This provides some additional evidence that helps support the existence of developmental stages of inductive reasoning. However, this result should be carefully interpreted, and future studies should employ a more balanced sample, from childhood to adulthood.

Study II: Refining the IRDT and investigating its construct/congruent validity

Study 2 aims to modify some items of the IRDT, based on the results from the first study, and, using Rasch analysis, assess its new scale structure, verifying whether the previously predicted orders and gaps, as well as the scale’s reliability and unidimensionality.

Part I: Instrument improvement

From the results of Study I, we’ve modified some items of the IRDT. Basically, the modifications can be synthesized as follows. From the original eight Pre-operational items, those demanding high perceptual discrimination were excluded, due to close similarities and low graphical clues (such as Q and O, etc), except one. We left one item to verify whether it still has more difficulties than the other Pre-operational items. The others were all modified in order to obtain items with easily discriminative options, such as "R F F F F" (Item Pre-op/sr3) and "H H L H H" (Item Pre-op/sr8). At the Primary order we removed those items constructed based on Rule 1, i.e. with no jump in the letters’ sequence, except for the option that is the exception and therefore is correctly supposed to be chosen by the participants because it does not follow the rule. Finally, the last change in the instrument occurred with the Abstract items, more precisely in the tables where the coordination of Concrete sequences are displayed. Instead of having a specific alphabetic letter in each row, and a specific Greek letter in each column, forming a code composed by two symbols for each cell that contains a coordination of two Concrete sequences, the table was modified to contain only one symbol (Greek letter) per cell. Moreover, the Abstract items are now formed by options that are spread throughout the table, so the participant needs to locate each one, and try to figure out which has a coordination rule that differs from the other 4 options. In the first version of the IRDT, the Abstract items’ options were organized in each row. Also, the “plus” (+) symbol that mediated the coordination of the two Concrete sequences was taken out. The other two orders’ items remained the same, since they demand the coordination of actions from the previous adjacent OHC. In sum, we’ve remodeled the items within each order, focusing on its vertical complexity. Our hypothesis is that this “verticalization” provides a better stage identification, with visual clusters of items and gaps between adjacent OHC more clearly defined.

![Figure 13](image_url)
METHOD

Participants

In Study 2, the revised IRDT were administered to a convenience sample composed of 188 Brazilian people (42.3% men, 57.7% women) aged between 6 to 65 years (M = 21.45, SD = 14.31). The sample, again, was intentionally broad and had a distribution of 34.4% from 6 to 12 years, 13.4% from 13 to 15 years, 7.5% from 16 to 21 years, and 44.6% older than 21 years. All the participants were from the city of Belo Horizonte, state of Minas Gerais.

Procedure

The data were collect by the first author and by twenty five Psychology undergraduate students, enrolled in a second semester Cognitive Development class, who were trained to administer the instrument properly. The author first administered the instrument to the undergraduate students (and those which data are actually being used in this analysis). Each undergraduate student had to administer the IRDT to different people from 6 to 65 years old. Participation was voluntary. The potential participants had the purpose of the study explained to them. They were informed that their answers would be kept confidential, and that all procedures guaranteeing the privacy of their results would be adopted. They signed a inform consent, according to the guidelines of the Ethical Committee of the Universidade Federal de Minas Gerais, Brazil.

Data analysis

The same data analytic process presented in Study 1 was adopted here. To assess the new scale structure of the IRDT, verifying if it presents the predicted orders and gaps, as well as its reliability and unidimensionality, we’ve employed the dichotomous Rasch model. To verify if the differences between the mean difficulty of items from order k and the mean difficulty of items from order k+1 are statistically significant, the one-sample t-test is used, with 95% confidence interval. The effect size is calculated using Cohen’s d.

RESULTS

The Rasch dichotomous model (Andrich, 1988; Rasch, 1960) was calculated using the software Winsteps (Linacre, 1999, 2011). From 48 items, only one was correctly responded to by all participants (Pre-op/SR8). The reliability for the full scale was .99, and its Infit mean was .94 (SD = .22; Max = 1.46; Min = .56). The person reliability was .95, which is estimated to indicate the degree to which a person’s response pattern conforms to the difficulty structure of the measure (Hibbard, Collins, Mahoney & Baker, 2009). The principal contrast showed that the raw variance explained by measures (modeled) was 74.8%, and that the unexplained variance in the first contrast (modeled) was 12.9%, suggesting that the instrument can be thought of as unidimensional, even though the variance explained by the first contrast is higher than 10%. We argue that the variance explained by measures (modeled) is high enough to sustain its unidimensionality.

The variable map (Figure 2) illustrates the scale for the 48 items of the IRDT with item difficulties (on the right) and person (student) measures (on the left) calibrated on the same scale. It’s visually possible to identify clear item clusters for almost all the orders, with a gap between them. However, two formal items, Form/AM6 and Form/AM8 had their scaled difficulties closer to the Systematic items, and one additional formal item, Form/AM3, had its scaled difficulty closer to the Abstract items. The only other difficulties were with the Pre-operational items, which were very spread out, but were nevertheless separated from the Primary items. Regarding the relative position of person (left) and item (right), the variable map shows the IRDT was an easy test for 28 participants (Mean ability = 7.86, SD = 0.87). The whole-sample mean ability was 1.15 with standard deviation of 3.40 logits (see Figure 15).

The one-sample t-test, with 95% confidence interval, shows that the comparisons between Pre-operational and Primary, Primary and Concrete, Concrete and Abstract, Abstract and Formal, and between Formal and Systematic were significant. Moreover, the effect size d' and r were large (see Table 3).

DISCUSSION

The evidence shows that modifying the IRDT, in order to eliminate some sources of horizontal complexity, produced an item structure closer to what was expected when constructing an instrument ac-

Figure 14. Variable Map showing the IRDT’s items
According to the MHC and using the strategies presented in the introduction (see Figure 1). In each OHC, the items are grouped forming a visual cluster, and presenting a gap in relation to the adjacent orders. Two Formal items had difficulties higher than expected (Form/AM6 and Form/AM8) and one was less difficult than predicted. However, this small deviation does not interfere with the spacing of its Rasch scores in relation to the adjacent orders of hierarchical complexity. The Pre-operational items have its scaled difficulties somewhat scattered through the less difficult end of the scale, an unexpected result to some extent, since the items were modified to contain stimuli that were expected to be easily discriminated (having many graphical clues). However, it can be speculated that the differences in difficulty of these items are due to factors other than the nature of each stimulus’ contribution to the increase in its horizontal complexity. In any case, the item Pre-op/SR4 presents a difficulty at least 1.26 logits higher than the other Pre-operational items. This result was expected, since the Pre-op/SR4 (“U U V U U”) is the same in both versions of the IRDT, and presents options graphically close to each other, demanding a higher amount of perceptual discrimination.

Regarding the data’s fit to the model, the modified version of the IRDT produced a better Infit mean of the items (.94), representing an increase of .06 over the items’ Infit of the first version (.88). The percentage of variance explained by the measures also increased from 70.6 with the previous version to 74.8 with the new one. It can be speculated that when we eliminated part of the horizontal complexity of the items, the amount of variance explained by the unidimensional measure increased. So, the “verticalization” process seems to contribute to the measure, not only in terms of the theory behind the items, i.e. the Model of Hierarchical Complexity, and by consequence the expected item structure, but also in terms of the adjustment of the items to the model and to the amount of variance explained.

Now that the item structure is closer to the expected (Figure 1), and the items’ fits are more adequate, it seems to be relevant to coordinate the Rasch metrics and the Orders of Hierarchical Complexity in a mathematical fashion, to obtain a score representing stage of performance. There is no direct way to obtain a person score that represents stage of performance from the estimates obtained through the Rasch Dichotomous model. This seems to be a dilemma, mainly because there is a difference in formal measurement theory terms between the OHC and the Rasch scores. The former is an analytic measure represented in an ordinal scale, while the latter are an empirical conjoint-interval measure. But, there’s a way to calculate stage of performance from the Rasch estimates. It can be calculated only because the items have the properties previously expected, i.e. they form clusters or groups within each OHC, present significant gaps with higher effect size between adjacent orders, and have adequate fit to the Rasch model. So, meeting these conditions, one can apply the below formula:

\[
\varphi_j = \beta_j - \frac{X_k}{X_{k+1} - X_k} + OHC_k
\]

where \(\varphi_j\) is the stage of performance of person \(j\), \(\beta\) is the Rasch score of that person, \(X_k\) is the mean difficulty of items on order \(k\), \(X_{k+1}\) is the mean difficulty of items on the next adjacent order, and \(OHC_k\) is the number that represents the order of hierarchical complexity \(k\). For computing the stage scores of people whose ability lies on the highest order measured, one needs to leave the denominator as \(X_k\). After computing the stage of performance for each person, it is possible to verify how well the stage scores regress on the order of hierarchical complexity of the items. Figure 16 shows the linear regression. As can be seen, the Order of Hierarchical Complexity of an item predicted the mean performance on that item with an \(R^2\) of 0.97 (see Figure 16).

**CONCLUSION**

In line with previous researches (Bond & Fox, 2001; Commons et al., 2008; Dawson, 2000, 2002; Dawson, Xie, & Wilson, 2003; Dawson-Tunik, 2004; Dawson-Tunik, Commons, Wilson & Fischer, 2005), the current study adds supportive evidence for developmental stages using modern quantitative methods and a specific test design provided by the model of Hierarchical Complexity and by


Figure 16. Regression of Stage Scores on Order of Hierarchical Complexity. The older version of the MHC stage numbers was used here. In the revised version, the stage numbers go up by one.

REFERENCES


J. Léna (Eds.), The educated brain: Essays in neuroeducation (pp. 127-150). New York, NY US: Cambridge University Press.


Richards, and C. Armon (Eds.), *Beyond formal operations: Late adolescent and adult cognitive development*, Vol 1, (pp. 43-73). New York: Praeger.


Linacre J. M. (2002). What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions*, 16 (2), 878


### APPENDIX A

**Description of the IRDT demands by OHC**

*Note.* The older version of the MHC stage numbers was used here. In the revised version, the stage numbers go up by one.

<table>
<thead>
<tr>
<th>OHC</th>
<th>Name</th>
<th>What they do</th>
<th>How they do</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>preoperational</td>
<td>Make very simple logical inductions, from single stimulus.</td>
<td>Proceeds from the identification and analysis of a group of single (equal) letters to a conclusion about an individual letter.</td>
</tr>
<tr>
<td>7</td>
<td>primary</td>
<td>Simple logical induction, from coordinated stimulus.</td>
<td>Proceeds from the identification of the relation between two coordinated letters, to a conclusion about a specific coordinated pair of letters.</td>
</tr>
<tr>
<td>8</td>
<td>concrete</td>
<td>Logical induction from a system of mapped stimulus.</td>
<td>Proceeds from the analysis of X pair of coordinated letters, forming a system of relations within a single option, to a conclusion about a specific coordination of X pair of letters.</td>
</tr>
<tr>
<td>9</td>
<td>abstract</td>
<td>Logical induction carried out through the comparison of single abstract, general, class of systems.</td>
<td>Proceeds from the identification and comparison of variables out of finite classes, to a conclusion about a specific variable.</td>
</tr>
<tr>
<td>10</td>
<td>formal</td>
<td>Logical induction from the coordinated abstract, general, class of systems.</td>
<td>Proceeds from the identification of the relation between two coordinated abstract variables, to a conclusion about a specific coordinated pair of variables.</td>
</tr>
<tr>
<td>11</td>
<td>systematic</td>
<td>Logical induction from a system of mapped abstract, general, variables.</td>
<td>Proceeds from the analysis of X pair of coordinated abstract variables, forming a system of relations within a single option, to a conclusion about a specific coordination of X pair of abstract variables.</td>
</tr>
</tbody>
</table>
### APPENDIX B

Inductive Reasoning Developmental Test 2nd Version

#### Pre-operational Items

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td>R</td>
<td>U</td>
<td>Q</td>
<td>V</td>
<td>D</td>
<td>H</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td>F</td>
<td>V</td>
<td>Q</td>
<td>V</td>
<td>G</td>
<td>H</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>C</td>
<td>F</td>
<td>U</td>
<td>Q</td>
<td>S</td>
<td>D</td>
<td>L</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>B</td>
<td>Q</td>
<td>U</td>
<td>Q</td>
<td>V</td>
<td>D</td>
<td>H</td>
</tr>
</tbody>
</table>

#### Primary Items

<table>
<thead>
<tr>
<th></th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>WX</td>
<td>IJ</td>
<td>TU</td>
<td>PQ</td>
<td>XY</td>
<td>ST</td>
<td>JK</td>
<td>GH</td>
</tr>
<tr>
<td></td>
<td>KL</td>
<td>RT</td>
<td>HI</td>
<td>NO</td>
<td>AB</td>
<td>CD</td>
<td>DE</td>
<td>XZ</td>
</tr>
<tr>
<td></td>
<td>ST</td>
<td>CD</td>
<td>QR</td>
<td>GI</td>
<td>TU</td>
<td>DF</td>
<td>OP</td>
<td>LM</td>
</tr>
<tr>
<td></td>
<td>PR</td>
<td>UV</td>
<td>JL</td>
<td>RS</td>
<td>EF</td>
<td>OP</td>
<td>NP</td>
<td>KL</td>
</tr>
</tbody>
</table>

#### Concrete Items

<table>
<thead>
<tr>
<th></th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>22</th>
<th>23</th>
<th>24</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NOPR</td>
<td>PQRT</td>
<td>HIJL</td>
<td>JKLN</td>
<td>OQEST</td>
<td>RTVW</td>
<td>IKMN</td>
<td>GIKL</td>
</tr>
<tr>
<td></td>
<td>IJKM</td>
<td>LMNP</td>
<td>TUWX</td>
<td>BCDF</td>
<td>DFHI</td>
<td>ACEF</td>
<td>LPNO</td>
<td>FHJK</td>
</tr>
<tr>
<td></td>
<td>UVXY</td>
<td>GHIK</td>
<td>RSTV</td>
<td>PQST</td>
<td>MOQR</td>
<td>BDEG</td>
<td>NLPQ</td>
<td>CEGH</td>
</tr>
<tr>
<td></td>
<td>MNOQ</td>
<td>VWXZ</td>
<td>OPQS</td>
<td>CDEG</td>
<td>EGHJ</td>
<td>CEGH</td>
<td>RTVW</td>
<td>CEGH</td>
</tr>
</tbody>
</table>

#### Reference Table

|   |   |   |   |   |
|---|---|---|---|
| Ж | Ю | ф | э |
| FGIKO | QST | OPRTDFH | IJLNPRI | EFHJTVXY |
| μ | ν | σ | ι |
| QRTVMOQR | STVXIKMN | KLMNOUWX | CDFHNRIS | GHIJLPRTU |
| Ω | Σ | Λ | Ψ |
| LMOQEGIJ | BCEGLJNO | MNPQIKL | JKMNOUWY | KLNPEI |
| θ | ξ | Π | A |
| UTVXKLNP | QSTVACEF | OQTBDGF | FHIKRTVW | HJMKGKL |
| α | τ | θ | β |
| OPQTCEGH | JLMOPRTU | UWXQSU | CEFHNNPS | HKMDFHI |
| Γ | ι | ε | ζ |
| KMNPQGIKL | EGHJGHI | QSTVMOQR | TVYKMP | VGSFUSW |
| Ψ | Ψ | Ψ | Ψ |
| CDGHUVWZ | KLOPEFGJ | CDGHUWHY | LMPQDGIE | QRUVMINOR |
| Β | Ψ | ι | ι |
| TUXYUKN | OPSTFGK | HILMNOPS | ABEFBCDG | UYVJQKMO |
### Abstract Items

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>Ж</td>
<td>Σ</td>
<td>Ф</td>
<td>Ε</td>
<td>η</td>
</tr>
<tr>
<td>26</td>
<td>μ</td>
<td>π</td>
<td>σ</td>
<td>Π</td>
<td>Ι</td>
</tr>
<tr>
<td>27</td>
<td>Ω</td>
<td>Σ</td>
<td>Δ</td>
<td>Ψ</td>
<td>Μ</td>
</tr>
<tr>
<td>28</td>
<td>Θ</td>
<td>Ξ</td>
<td>Π</td>
<td>Ψ</td>
<td>α</td>
</tr>
<tr>
<td>29</td>
<td>αε</td>
<td>Τ</td>
<td>ρ</td>
<td>β</td>
<td>δ</td>
</tr>
<tr>
<td>30</td>
<td>Σ</td>
<td>π</td>
<td>ε</td>
<td>Ζ</td>
<td>λ</td>
</tr>
<tr>
<td>31</td>
<td>Ѳ</td>
<td>ρ</td>
<td>ς</td>
<td>й</td>
<td>γ</td>
</tr>
<tr>
<td>32</td>
<td>Б</td>
<td>Υ</td>
<td>θ</td>
<td>Τ</td>
<td>η</td>
</tr>
</tbody>
</table>

### Formal Items

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>33</td>
<td>ЮЪ</td>
<td>λαε</td>
<td>πδ</td>
<td>Σε</td>
<td>ημ</td>
</tr>
<tr>
<td>34</td>
<td>Ζъ</td>
<td>Σλ</td>
<td>Ψα</td>
<td>Ωδ</td>
<td>αΞ</td>
</tr>
<tr>
<td>35</td>
<td>ЮΩ</td>
<td>μλ</td>
<td>οΓ</td>
<td>ΙΞ</td>
<td>ΗΨ</td>
</tr>
<tr>
<td>36</td>
<td>Ψγ</td>
<td>ρτ</td>
<td>οΞλ</td>
<td>οΤ</td>
<td>Τρ</td>
</tr>
<tr>
<td>37</td>
<td>ΨЪ</td>
<td>δБ</td>
<td>λ‡</td>
<td>Ξε</td>
<td>ΞΨ</td>
</tr>
<tr>
<td>38</td>
<td>λπ</td>
<td>αΨ</td>
<td>ρμ</td>
<td>Βσ</td>
<td>δΠ</td>
</tr>
<tr>
<td>39</td>
<td>έε</td>
<td>αηΟ</td>
<td>Ψη</td>
<td>ΞΩ</td>
<td>εΙ</td>
</tr>
<tr>
<td>40</td>
<td>ΤΞ</td>
<td>ПЮ</td>
<td>δΣ</td>
<td>ζΛ</td>
<td>εΨ</td>
</tr>
</tbody>
</table>

### Systematic Items

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>41</td>
<td>ЮσЪΞ</td>
<td>Δαε</td>
<td>ΙΞ</td>
<td>ЮЪΞ</td>
<td>ΜλΣξ</td>
</tr>
<tr>
<td>42</td>
<td>ΨαΙΞ</td>
<td>ΔΙФμ</td>
<td>ΣΨΔΤ</td>
<td>παφ</td>
<td>ΗΨΗЪ</td>
</tr>
<tr>
<td>43</td>
<td>ξσΣΠ</td>
<td>ΗδΩδ</td>
<td>ЮζΩλ</td>
<td>ΨδΩЪ</td>
<td>μΞΣε</td>
</tr>
<tr>
<td>44</td>
<td>ПЮδμ</td>
<td>Ωσαε</td>
<td>δΣΞΩ</td>
<td>ΛπεΙ</td>
<td>ΒσαЮ</td>
</tr>
<tr>
<td>45</td>
<td>αΨΤΞ</td>
<td>ΛδЪЖ</td>
<td>δΣαεω</td>
<td>εΩΞη</td>
<td>ζπμα</td>
</tr>
<tr>
<td>46</td>
<td>ΨΣμα</td>
<td>ρЮΨΩ</td>
<td>ΠαΞμ</td>
<td>ТΗЪΗ</td>
<td>αλδ</td>
</tr>
<tr>
<td>47</td>
<td>ΨЮΞτ</td>
<td>ΤЮΟΤ</td>
<td>Χαφ</td>
<td>αφσα</td>
<td>ΕПΨь</td>
</tr>
<tr>
<td>48</td>
<td>δБαЮ</td>
<td>ΨΓЛ‡</td>
<td>ΡΤΞЮ</td>
<td>ТЪЪЪ</td>
<td>ΖЪΣΠ</td>
</tr>
</tbody>
</table>