International Journal of School and Cognitive Psychology

Kubinger, Int J Sch Cog Psychol 2017, 4:1
DOI: 10.4172/2469-9837.1000189

Research Article OMICS International

Adaptive Intelligence Diagnosticum (AID): An IRT-Based Intelligence Test-Battery Fulfilling Many Practitioners' Requests

Klaus D Kubinger*

Division for Assessment and Applied Psychometrics, University of Vienna, Vienna, Austria

***Corresponding author:** Klaus D Kubinger, Division for Assessment and Apllied Psychometrics, Faculty of Psychology, University of Vienna, Vienna/Austria, Vienna, Austria, Tel: 0043/1/4277-478-50; E-mail: klaus.kubinger@univie.ac.at

Rec date: Oct 25, 2016; **Acc date:** Nov 14, 2016; **Pub date:** Nov 14, 2016

Abstract

In this paper the intelligence test-battery AID (Adaptive Intelligence Diagnosticum; e.g. Kubinger, 2017) is introduced as an extraordinary economical means for practical case consulting of pertinent school and educational psychological problems. High measurement accuracy is provided due to adaptive testing, where almost only half the number of items is needed as compared to conventional tests. For this, a profile interpretation of the subtests' results is actually serious, as the incorporated measurement errors are rather small, which entails that overlapping confidence intervals of the subtests' results do not occur very often. All this is based on high psychometric standards, that is, Item Response theory (IRT). Furthermore, AID is attractive for practitioners who are aware that sometimes some "partial performance weakness" may be (additionally) causing the respective problem; there are several optional tests involved in a process-oriented neuro-functional hierarchy of basic abilities.

Keywords: Enhancement-oriented psychological assessment; Adaptive testing; Accuracy of measurement; Profile interpretation; AID (Adaptive Intelligence Diagnosticum)

Introduction

Intelligence tests are without a doubt still a very important means in practical case consulting for many school and educational psychological problems. Numerous published single case studies prove their significance [1]. Admittedly, in the long run, it seems preferable to call them "tests for ascertaining (as much as possible) cognitive abilities" instead of intelligence tests; however, the term "intelligence" is accepted by lay persons and they know its meaning to a satisfactory extent. Intelligence, with regard to the test-battery we will deal with in the following, is hence defined: "Intelligence is the entirety of all cognitive conditions which are necessary to acquire knowledge and develop action competence" [2].

To assess "intelligence" comprehensively, a test-battery is needed with as many subtests as necessary for analyzing almost all the cognitive conditions mentioned. A single facet, for instance even reasoning, does not suffice for most of the pertinent school and educational psychological problems. That is, in contrast to the simple ascertainment of the IQ, case consulting for these problems demands an enhancement-oriented psychological assessment: when a diagnosis is established, intervention measures should be obviously evident, as well. Such an approach has to aim for a differentiated profile interpretation of many subtest results.

In this context, test-batteries conceptualized for individual administration, in which the test-administrator tests a single examinee, are to be favoured in comparison to a group administrable test-battery. For instance, Kubinger, et al. [3] discuss the psycho-diagnostic added value of individual settings in depth. Pertinent tests use materials which only cover certain specific problem-solving cognitions; and they deliver valuable information of a child's attitude to work by observing its problem-solving behavior-apart from the fact that such tests do not need a multiple-choice answer format, which only delivers the information of recognizing the item's solution rather than of producing the solution, challenges young examinees' memory (and strategical competence), and offers the possibility for lucky guessing; and such tests are not forced to be rushed due to time restrictions, which is necessary in group administration.

Evidently, there are some test-batteries at a practitioner's disposal which can be utilized for this need. In this paper, we introduce a new one-though its German original issues from 1985, this test-battery most importantly complements and brings added values to the market. This is because, 1) it verifiably holds essential psychometric requirements, 2) it is extraordinarily economical regarding the relation of number of administered items and accuracy of measurement, 3) it also serves for the assessment of basic neuro-functional disorders, or to say, of "partial performance weaknesses". To summarize: it fulfils many practitioners' requests.

Materials and Methods

The test-battery AID (Adaptive Intelligence Diagnosticum) aims to measure most of the realizable aspects/dimensions of cognitive abilities, complex as well as basic ones. It can be administered to children and adolescents of the age of 6-16, and even to older people. The German up-to-date version was written by Kubinger and Holocher-Ertl [4], the English version by Kubinger [2]; an additional Japanese and an additional Russian version is forthcoming.

Not only for psychometricians, but also for practitioners, of high interest is the fact that-if presupposed by the respective scoring rule-the Rasch model holds empirically proven for every subtest. That is, the test-battery AID does guarantee fair scoring (cf. Kubinger, 2005, p. 378: "if the number of solved items is to be considered a sufficient statistic, then all items of the test must fit the Rasch model") [5]. No differential item functioning takes place insofar that no item in a sub-group of the population handicaps or benefits examinees systematically. Be aware that any competing test-battery does not stand

this test or, at least, does not provide respective proof. Also, be aware that when consulting is based on almost dubious measurements, the practitioner always risks the possibility of subsequent lawsuits.

Based on this psychometric high standard, the AID makes adaptive testing possible. Basically, adaptive testing means some optimal item selection on the basis of the examinee's performance on preceding items. As a consequence, hardly any items occur for any examinee that are too difficult and therefore maybe frustrating; and hardly any items occur which are too easy for him/her and therefore maybe de-motivating. Apart from these motivational aspects, adaptive testing serves for high measurement accuracy as non-informative items are not needed to be administrated; coming to the point: using only half of the number of items in comparison to conventional, that is fixed item testing, the same accuracy of measurement can be gained by adaptive testing. We give an illustration below. However, it is necessary that the item-pool of the test in question suits the frame of Item Response theory (IRT), as the looked-for ability parameter is only calculable by these means and comparable with examinees who had (completely) different items administered (see for an introduction into adaptive testing e.g. Kubinger, 2016) [6]. As indicated above, this matter is given for the AID: all but one concerned subtest conform to the Rasch model, the remaining one to some generalization of the Rasch model.

Exemplarily, the following branched testing design is used for most of AID's subtests (Figure 1). Starting with a first subset of five items, appropriate to the age, a second and subsequently a third subset of five items is administered; both of the latter depend on the preceding score of the examinee. If he/she solves one item at the most, a subset one level below is recommended. If at least four items are solved then a subset one level higher is advised. Only if two or three of the five items are solved is a subset at an equivalent level administered. This design guarantees that every examinee is tested with almost optimal items.

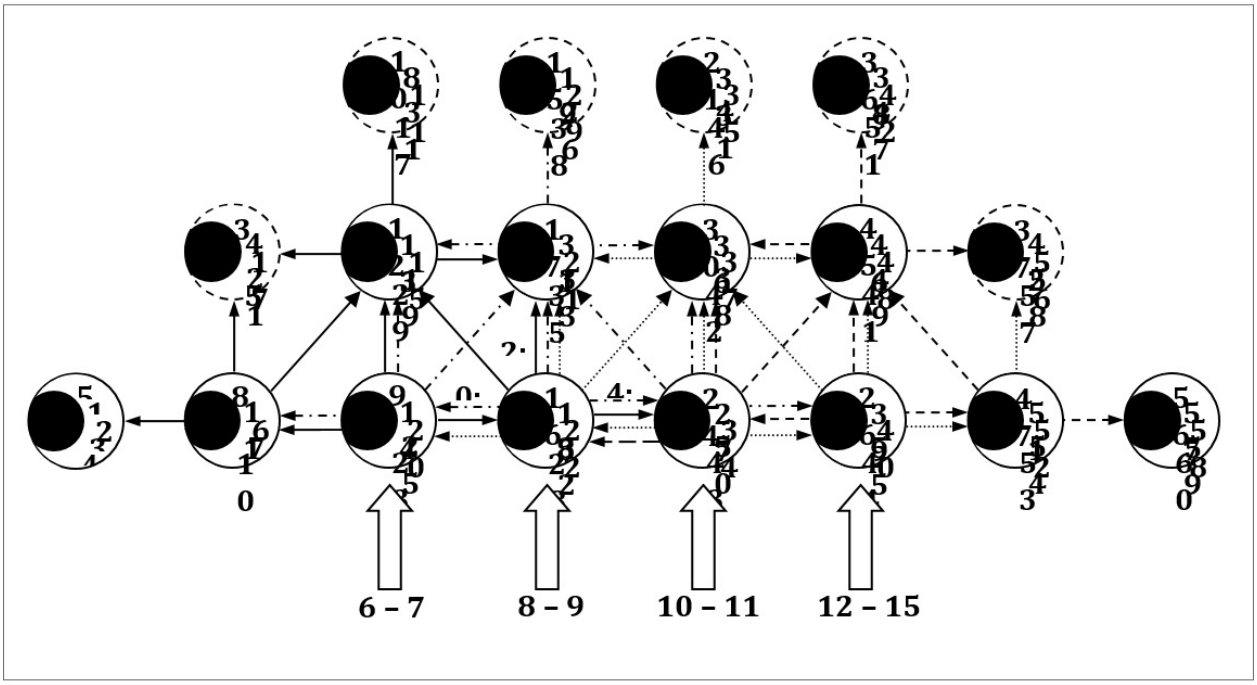

Figure 1: The branched testing design exemplary for AID's subtest everyday knowledge.

The circles represent different groups of items, arranged from left to right according to their averaged item difficulty. Each item group contains five items; the code number of each corresponds (roughly) to the rank order of all the items' difficulties. Based on the examinee's age, a certain starting group is to be administered. The next item groups are then chosen according to the examinee's performances, that is, according to the number of correctly answered items in an item group. If the examinee solves no items or only a single item, an easier item group is to be administered; if he/she solves four or five, a more difficult one; only if the examinee solves two or three items, then an almost equally difficult item group is to be administered next. By default, three item groups are to be administered for each examinee. Given the testee is 8 or 9 years old, the Figure shows in detail which item group combinations could occur, depending on previous performances. As branched testing terminates after the third item group, dashed-lined groups also consist of items from the solid-lined groups; that is, the whole item-pool consists of 60 items.

Figure 2 illustrates the superiority of AID's branched testing with respect to the number of items and accuracy of measurement, respectively. Given is the (expectation of the) standard error of the looked-for ability parameter estimation. Five cases are presented: Firstly, for the branched testing design with 15 items according to Figure 1, that being four different curves, one for each of the four age-homogeneous groups-bear in mind that only the lower "infolding curve" of the standard error of estimation is of practical interest. Then, the standard error estimation is presented if all 60 items would be administered (conventionally), which is not really of any practical use for practitioners, however does offer a good impression of the superiority of adaptive testing. Thirdly, the respective curve if 30 items with rather medium item difficulty were to be administered

(conventionally)-this case is almost in accordance with the use of any familiar test-batteries' subtests. Finally, there are two lines; the first of them, the solid line, refers to 60 theoretically ideal items-these are not really but rather hypothetically existent and endure probabilities of consistently. 50 to be solved, the second one, the dashed line, refers to 15 theoretically ideal items.

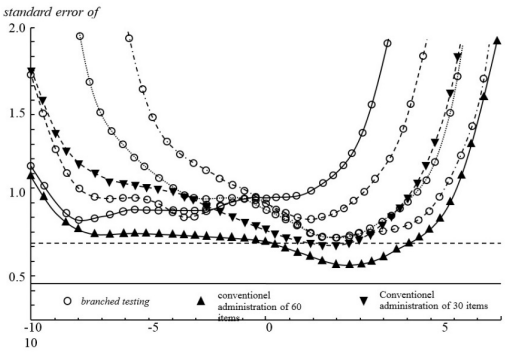

Figure 2: The for three different administrations: branched testing each examinee with 15 items according to the design in Figure 1 (four starting group-specific curves), conventional testing with all 60 items, and conventional testing with 30 items predominantly with medium difficulties.

Comparing the lower "infolding curve" with the case of administering all 60 items of the pool to every examinee (and more in the case of 30 items predominantly with medium difficulties) emphasizes a proper accuracy of measurement of 15 adaptively administered items. Only in a quite limited range of medium abilities-which is of minor importance for assessment practice-is the standard error of estimation definitely larger than when, unrealistically, using the whole item pool of 60 items.

As a consequence of acceptable low errors of measurement due to very high measurement accuracy, i.e. small confidence intervals for each subtest result, a profile interpretation actually applies. Bear in mind that conventional test-batteries with only a measurement accuracy as illustrated in Figure 2 by the curve of 30 medium difficult items, very often lead to a profile with overlapping confidence intervals, that is, there are hardly reliable differences in an examinee's test results, but rather a lot of them disclose the same extent of ability when the error of measurement is taken into account.

Contents

The AID's theoretical basis is a simply pragmatic one: as already indicated, it tries to measure very many (complex as well as basal) abilities which all seem to be responsible for "intelligent" behavior. Every subtest has been conceptualized according to appropriate copyright operational definitions and rules of item generation. There are 14 subtests, with either "verbal-acoustic" or with "manual-visual – tasks:

Everyday knowledge refers to the ability to inform oneself about things that are commonly present nowadays.

Competence in realism means the ability to understand the functional properties of everyday objects.

Applied computing deals with the ability to solve every day numerical problems, primarily by reasoning, however not by using any high-level school numeracy techniques.

Social and material sequencing means the ability to understand and control the sequence of everyday actions, concerning social, as well as material conditions.

Immediately reproducing (numerical information) demonstrates the (verbal-acoustic) capacity of the short-term memory with respect to consecutively presented stimuli.

Producing synonyms refers to the ability to understand the concept of a word and the ability, respectively, to use an alternative word instead.

Coding and associating demonstrates speed of information processing and incidental learning ability.

Anticipating and combining (the parts of a cut up "Gestalt") establishes deductive reasoning abilities in regard to recognizing and composing an everyday object in its entity.

Verbal abstracting (the common functionalisms of things) refers to the ability of conceptualization.

Analyzing and synthesizing (abstract figures) refers to the ability to structurally decompose and rebuild a "Gestalt".

Social understanding and Material Reflection incorporates the comprehension of factual everyday connections and the knowledge of socially expected behavior.

Formal sequencing (of geometric figures) measures the ability to perceive regularities and logically consistent dependencies, and to utilize them appropriately.

Furthermore, there are five optional subtests, also either with "verbal-acoustic" or with "manual-visual" tasks:

Immediately reproducing (figural stimuli) demonstrates the capacity of serial information processing (within visuo-motoric domains).

Storing by repetition demonstrates the (lexical) capacity of memory after stimuli are presented twice in a row.

Learning and long-range memorizing acquires learning efficiency and capacity of long-term memory (within the domain of spatial perception).

Producing antonyms refers to the ability to understand the concept of a word by producing the antipode.

Recognition of (figural) structures refers to the ability to decompose complex abstract figures into given units.

Enriched with these optional subtests, the test-battery AID offers a survey for the case that specific neuro-functional (developmental, learning) disorders-or to say "partial performance weaknesses"-are suspected; thereby such a suspicion might even only arise after the application of the obligatory subtests. The respective systematization is process-oriented and determines a functional hierarchy: it distinguishes between perception, retrieval, and utilization. Perception applies to visual, acoustic, and proprioceptive activities; visual and acoustic modalities refer either to lexical or numeric tasks; visual modalities additionally to figural contents. There are the processes of classifying and discriminating (space localization included) on the one hand, and sequencing, on the other hand. Sequencing means the

ability to handle consecutive stimuli appropriately. Retrieval refers to the same modalities and contents; it only functions well if perception has already been mastered. That is, retrieval might not necessarily function well, even though perception does. Apart from modality and general contents, as well as apart from modality and a content specific disorder of retrieval, there may also be general or specific deficiencies in regard to sequencing. Utilization also involves the same modalities and contents, but frequently also concerns intermodal transfers with motor aspects-at least with regard to speaking. Bear in mind, that intermodal and motor-, particularly visuo-motor abilities can be compromised, even though perception and retrieval function well. For all this, according to this systematization, several patterns of the AID's test results in all the sub- and optional tests refer to specific "partial performance weaknesses".

Practitioners' Requests

We are confident that practitioners appreciate AID's potential of serious profile interpretation of the subtests' results due to the high measurement accuracy provided, when, by the way, rather fewer items are needed than with a conventional test. The latter finally means shorter test duration: administration of all 14 AID subtests rarely requires more than 75 minutes. Of course, the user's effort as concerns the calculation of the test results is minimized by an optional provided computing program [7]. At any rate, test administration according to the branched testing design in Figure 1 is handled rather easily.

Additionally to the recommended profile interpretation, the "partial performance weakness" assessment indicated above serves the profit of applying AID. However, there are other issues which fulfil many practitioners' requests on an intelligence test-battery when dealing with pertinent school and educational psychological problems:

There is a supplemental sheet for categorizing the examinee's "attitude to work" which acts as a qualitative assessment of his/her working and contact behavior.

There are parallel forms for eight subtests which are of use if test repetition of an examinee is necessary.

The availability of instructions for non-verbal administration for six sub- and three optional tests enriches the field of possible application.

The items are as far as possible globalized, that is, all existing and proposed versions do not and will not suffer from differential item functioning with respect to the cultures concerned.

It is possible to administer the test-battery by default as suggested, but without using the age-specific start group of items; this may be of relevance if it is suspected that the testee suffers from low giftedness or the other way round, has a high ability.

Recent results show that the test-battery AID is also applicable for examinees older than 16, the norms of which do not differ enormously.

Above all it should be added, that most recently even a software is at the practitioners' disposal, which realizes a computer-assisted AID administration and applies an even more effective adaptive testing procedure, that is, after each item, not only after e.g. five items, the optimal next one is selected by the computer; by this means, instead of e.g. 15 items on average, just 11 or 12 suffice [8].

Conclusions

The test-battery AID is of particular use when practitioners' look for an instrument which directly indicates specific preventative or rehabilitative treatment of pertinent school and educational psychological problems, that is, when they aim for enhancement-oriented psychological assessment. This is true, above all, when they prefer to do their job economically regarding the relation of the number of administered items to the accuracy of measurement. Furthermore, this test-battery is attractive for them, if they are aware that sometimes a "partial performance weakness" may be (additionally) causing the respective problem, which, however, can only be suspected during the assessment process.

References

1. Kubinger KD, Holocher-Ertl S (2012) Fallbuch AID (case studies of AID). Hogrefe, Göttingen.
2. Kubinger KD (2017) Adaptive intelligence diagnosticum (AID). Hogrefe, Göttingen & Oxford.
3. Kubinger KD, Deimann P, Kastner-Koller U (2012) The psycho-diagnostic added value of individual settings. In: Kubinger KD, Holocher-Ertl S (Eds.) Fallbuch AID (case studies of AID) 21-28.
4. Kubinger KD, Holocher-Ertl, S (2014) Adaptives Intelligenz Diagnostikum-Version 3.1 (AID 3) (Adaptive Intelligence Diagnosticum, AID 3). Beltz, Göttingen.
5. Kubinger KD (2005) Psychological Test Calibration using the Rasch Model-Some Critical Suggestions on Traditional Approaches. International Journal of Testing 5: 377-394.
6. Kubinger KD (2016) Adaptive testing. In: Schweizer K, DiStefano C (Eds.), Principles and methods of test construction Hogrefe, Göttingen 104-119.
7. Spohn F (2014) Evaluation program AID_3_Score for AID 3 [Test score computation program AID_Score]. Hogrefe, Göttingen.
8. Kubinger KD, Spohn F (2017) Computer-aided specification of the AID 3 according to the principle of tailored testing. Manual for the software AID_3_tailored. (Computerassisted tailored testing of AID. Manual of the software AID_tailored). Hogrefe, Göttingen.