



International Arab Baccalaureate



IAB 2010 Pilot Summative Report

July 2010

BACKGROUND

The International Arab Baccalaureate (IAB) is a comprehensive educational program that Educational Research Center (ERC) has launched for implementation in the academic year 2010-2011 all over the Arab World. It came about after many years of preparation, including two years of piloting at a number of schools in four Arab countries. This summative report presents major results of the IAB pilot conducted in February-March 2010 in Egypt, Jordan, KSA, and Lebanon.

IAB is designed in the framework of *Profile Shaping Education* (PSE), a novel research-based educational framework developed by Prof. Ibrahim Halloun and colleagues, under the auspices of ERC. PSE empowers students with the *profile* needed for success in various aspects of modern life. The profile consists of an epistemic dimension and a cognitive dimension. The *epistemic* dimension consists of a coherent repertoire of conceptions (concepts, laws, models and other conceptual entities) drawn from the formal disciplines which various curricula are about. The *cognitive* dimension includes generic habits of mind, i.e., skills and dispositions (values, attitudes, affects, etc.), that are critical for meaningful learning of course materials, as well as for lifelong learning and success in modern life. Cross-disciplinary habits of mind are the most critical in these respects, especially for realizing the big picture within and across disciplines. Such habits of mind empower students to be *paradigmatic* thinkers, as well as *productive*, *proactive* and *principled* citizens (whence the name “4-P Profile”).



PSE focuses on *patterns* in both the physical world and the mental realm of human beings, especially patterns of thought that are shared by various successful professionals. These patterns are best manifested in *systems* of well-defined scope and structure at the epistemic or content level, and in well-defined categories of habits of mind, or skills and dispositions, at the cognitive level. The epistemic and cognitive dimensions are systematically ascertained in the framework of PSE in accordance with a novel taxonomy of learning outcomes.

At the epistemic level, the taxonomy concentrates on the scope and structure of physical or conceptual systems. The *scope* dimension specifies the *domain* of a system (what pattern the system represents in either the physical world or conceptual realm) and its *function* (what the system is good for, and under what conditions). The *structure* dimension specifies the *composition* of the system (what primary entities the system consists of, and what are their salient properties), its *internal structure* (how these elements and their properties are related to each other within the system), and its *external structure* (how the system relates to its environment and/or other systems within and outside the confinement of its paradigm).

At the cognitive level, the taxonomy concentrates on seven process or skill categories and six disposition categories. The process categories include: analysis, criterial reasoning, relational reasoning, critical reasoning, logical reasoning, technical dexterity and communication dexterity (Figure 1). The disposition categories include affects, attitudes, morals, ethics, values and beliefs. Some dispositions, like affects and attitudes, are *inward* personality traits that govern one’s own conduct (affects) or position toward, and thus conduct with, others or things (attitudes). Other dispositions, like morals and ethics, are primarily customs and rules of conduct *set by others*, mainly a given society (morals) or profession (ethics). Some other dispositions, like values and beliefs, result from a mix of both intrinsic traits and external controls.

The 7-category cognitive taxonomy includes:

Analysis, which includes exploratory processes allowing the description or explanation (cause identification) of the state or change of state of a given system (or phenomenon) and inferential processes allowing, among others, the prediction (or post-diction) of the state of the system under certain conditions. In all these respects, analysis may be either exhaustive or discriminative. Discriminative analysis is conducted to tease out primary or salient features (entities and their properties) from secondary or irrelevant features, whereas exhaustive analysis results in identifying all features without any distinction.

Criteria reasoning, which includes all sorts of criteria-based judgment and evaluation, like comparison, contrast, classification, analogical reasoning, and pattern recognition, as well as estimation and measurement, all done with special attention to reliability, consistency, objectivity and precision.

Relational reasoning, which includes processes establishing viable (valid and reliable, coherent and consistent) relationships between different features, including syntactical connections, internal or cohesive structure of a system (connecting its features) or its external structure (connecting the system to its environment), correlation, functional relation, synthesis, extrapolation, transfer.

Critical reasoning, which includes processes of reflective thinking, evaluation of claims and evidence, corroboration of claims and hypotheses, question formulation, problem detection and formulation, challenge anticipation, skepticism and questioning “facts”, all done with special attention to objectivity and precision.

Logical reasoning, which includes processes of evidence-based argument and corroboration, justification, proof, hypothesis formulation, assumptions making, conjecturing, adduction, induction, deduction, generalization, metaphorical reasoning, esthetical reasoning, insight.

Technical dexterity, which is about efficient and constructive use of computers, ICT media and all sorts of technical devices that are particularly important in education.

Representation dexterity and communication fluency, which include verbal and symbolic expression, graphic and geometric depiction, kinesthetic expression, coordination of various expressions and depictions, semantic processes of interpretation and sense making, all done with eloquence, clarity, objectivity and precision.

It is important to note at this point that there is no particular cognitive hierarchy among the various categories. However, a certain hierarchy may be identified within each category that depends on the variation of complexity of, and cognitive demands imposed by, each habit of mind within a given category. For example, within the category of analysis, we may distinguish between exploratory analysis and inferential analysis. Exploratory analysis is about describing or explaining a particular state of a given system, as it exists at given point of space and time. Inferential analysis is about making inferences about the system in question beyond that particular state, e.g., predicting how the system may evolve in the future under certain conditions, or post-dicting how the system evolved in the past before it got to the state in question. One can readily realize that inferential analysis comes at a higher cognitive level than exploratory analysis, and that explanatory analysis (identifying salient causes of the conservation or change of state of a system) comes at a higher level than descriptive analysis (identifying primary features of a given state).

A particular cognitive hierarchy is defined in PSE that relates to the gradual construction and deployment of a given system along the dimensions defined in the profiling schema (Figure 4). Accordingly, any person may progressively “know” (or “learn” about) a given system, and develop and deploy any conception or habit of mind, in four consecutive stages. These are in order:

1. *Initiation (primitive learning)*, when a learner is simply aware that the system exists, but knows nothing or a little about its scope and structure, and is still incapable of successfully deploying related conceptions and habits of mind in any situation.
2. *Gestation (rote learning)*, when the learner develops partial knowledge about the scope and structure of the system, and is capable of deploying certain related conceptions and habits of mind, exclusively in the context of the system in question when encountered in familiar situations.
3. *Replication (reproductive learning)*, when the learner develops satisfactory knowledge about the scope and structure of the system, and is capable of deploying related conceptions and habits of mind, exclusively in the context of the system in question when encountered in familiar situations and new, but mostly similar, situations.
4. *Innovation (productive or meaningful learning)*, when the learner develops comprehensive knowledge about the scope and structure of the system, and is capable of creatively deploying this knowledge, especially corresponding habits of mind, within the context of the same and other systems encountered in novel, unfamiliar situations.

It is also important to note here that the habits of mind distinguished above are normally gradually developed and deployed not individually but together in various combinations that may be classified in three categories: core-engagement, eco-engagement, and meta-engagement.

Core-engagement, which puts together processes and dispositions from the above taxonomy of habits of mind, in the purpose of looking at the big picture of things within a given field, and designing and carrying out appropriate plans for systemic and system-based thinking (modeling included), problem solving and experimenting with, and regulating (controlling or changing) existing situations in that field. Ultimately, one will bring about creative and innovative ideas and products about the field, at school or in the workplace, as well as in everyday life.

Eco-engagement, which includes self-management as well as interaction with others, especially peers (teamwork included), and the environment, decision making, and crisis management, all done with integrity, fairness, tolerance, empathy, respect of diversity, open mindedness, while upholding commitment, dedication, perseverance, curiosity, adaptability, and assuming responsibility, accountability, leadership.

Meta-engagement, which includes auditory, visual, and/or kinesthetic assimilation and adaptation of conceptions, and various conscious actions for the development of habits of mind, as well as various meta-cognitive controls that govern learning and, especially, learning how to learn.

Figure 1: PSE Cognitive Taxonomy (in Halloun, 2010: *From Modeling Schemata to the Profiling Schema: Modeling across the Curricula for Profile Shaping Education*. A Springer publication)

According to PSE, any educational activity, from curriculum design down to programs of study and means and methods of learning, instruction and assessment, should promote the development of the 4-P profile. The profile would be practically realized in the classroom in the form of epistemic and cognitive learning outcomes that need to be well-defined within the scope of well-chosen systems. A *learning outcome* is an aspect in student epistemic product and/or cognitive process exhibited in *measurable* ways in an assessment task administered to the student. The extent to which an outcome is reified (achieved) is determined by specific *indicators* with which are associated particular *scales*. An *epistemic outcome* is an outcome that tells what “information” a student actually “has”, and meaningfully “understands”, about the scope or structure of a system or conception. A *cognitive outcome* is an outcome that tells what a student “is” “like” (disposition) or is capable of “doing” (skill), mentally or physically, with respect to a given category in our cognitive taxonomy, or in relation to the student learning style.

IAB is gradually deployed at schools accredited by ERC in various parts of the Arab World, within the framework of existing curricula. It does not impose any rigid program of study, or any textbook, on any school. It works, instead, to infuse harmony within diversity, by focusing on common learning outcomes in core and optional fields from humanities, mathematics, science and technology. These outcomes concentrate on generic conceptions and habits of mind, which bring about the big picture within and across disciplines and efficiently realize the IAB student profile.

Deployment begins in 2010-2011 with Grade 10 students to cover, by 2012-2013, the three secondary school grades, and subsequently K-12. Various components of IAB are also gradually deployed throughout the years, while teacher capacity is being progressively built to master all aspects of PSE and IAB. The focus in IAB begins on authentic assessment, while the grounds are being prepared for the experiential learning approach that PSE calls for, and various other aspects of learning and instruction.

Educational research has been showing during the last two decades that a major problem with education is that assessment is often looked at as a major goal instead of a guide for learning and instruction, and that it subsequently drives students toward rote learning rather than meaningful learning of course materials. Most importantly, assessment is being blamed for not providing reliable measures of the actual student knowledge state, and that students often pass, and even excel on their exams only because they are capable of recalling certain information or routines for answering questions or solving problems, without necessarily understanding what these questions or problems are about. Because “what you measure is what you get”, and because through assessment, teachers are better prepared to begin to understand what PSE is all about, the focus in IAB has begun with assessment, at the pilot stage, beginning in May 2009, and at the formal deployment stage starting in the fall of 2010.

There were two major objectives behind the two-year pilot of IAB: (a) to ascertain whether the PSE framework is a viable common framework for all educational fields, particularly with respects to the novel taxonomy of learning outcomes at the cross-disciplinary boundaries, and (b) to find out whether IAB can actually be deployed within the framework of curricula currently in place at various schools and parts of the Arab World. The first pilot consisted of summative exams administered in May 2009 to a sample of Grade 12 students in four Arab countries, Egypt, Jordan, KSA, and Lebanon. That pilot report is available at: www.EducationalRC.org/IAB. The second pilot consisted of midterm summative exams administered in February-March 2010 to a sample of Grades 10, 11 and 12 students in the same four countries. The last pilot is the object of this summative report. Data and analysis herein reported are kept to the minimum necessary for an ordinary reader with limited background in educational research to be able to follow through.

METHOD

The 2010 pilot consisted of midterm summative exams administered to Grades 10, 11 and 12 students in the following fields: physics, chemistry, biology, mathematics, social studies (geography), Arabic, and a foreign language (English or French). Depending on their language of instruction, students took the science and mathematics exams in Arabic, English or French. Except for languages, all tests consisted of multiple-choice items. Each language exam consisted of a number of multiple-choice items in addition to one open item where students were asked to write about a particular topic.

Themes covered in each field are outlined in Figure 2. More details are available at: www.EducationalRC.org/IAB. Outcomes were set in each field in the context of the themes in question, with special attention on cross-disciplinarity. Epistemic outcomes in all fields were about the scope and structure of specific systems pertaining to the themes of Figure 2. Cognitive outcomes in all fields concentrated on five skill categories: analysis, critical reasoning, relational reasoning, logical reasoning, and communication dexterity.

Exams were designed to allow the pilot to answer the following two questions:

1. Does the taxonomy of epistemic and cognitive learning outcomes specified for IAB actually correspond to a coherent student profile? In other words, to what extent outcomes specified in various fields are actually cross-disciplinary, and how they correlate with each other?
2. To what extent IAB is viable for schools in various parts of the Arab World? Can it be readily deployed in those schools in the context of existing curricula?

While answering those questions, we also sought to ascertain the impact on student performance in the pilot of various demographic and curricular variables (learning style, teaching approach, student performance on regular school and state exams, etc). Unfortunately, we had access to data on a limited set of variables as shown in the following section.

Pilot exams were administered during two to three consecutive days at various schools during the period extending from the 21st of February to the 20th of March 2010. Students were provided with separate questionnaires and answer sheets for each exam. All answers were to be marked on specially designed OMR sheets (scannable bubble sheets for multiple-choice items). Students were asked to write the answers of the multiple choice questions on bubble sheets and finish their work within 100 minutes.

An IAB School Officer and science and humanity coordinators were appointed by the administration of each school to coordinate pilot matters with ERC, and oversee exam administration in accordance with written guidelines. These guidelines were set following a pilot orientation and preparation conference attended in December 2009 by these officers and other school administrators. Test administration went accordingly without any incident. All test booklets and answer sheets were returned on time to ERC. OMR sheets were scanned and open item manually graded on ERC premises. Each open item (in languages only) was graded by two separate graders, and scores consolidated by a third person (the IAB coordinator of the language in question). All data was then securely entered into the ERC data bank with strict confidentiality, and made ready for statistical analysis. No one has ever had, or will ever have access to the data except the two psychometricians in charge and the IAB chair.

In the following, we report on overall results pertaining to the two questions above without revealing particular data about any given country or school. Meanwhile, each school exclusively receives, along with this report, a copy of its own students' normalized scores.

Physics

1. Newtonian Mechanics and Fluid Mechanics
2. Electricity & Magnetism
3. Waves
4. Oscillations and EM Induction

Biology

1. Reproduction and Genetics
2. The functional characteristics of cells
3. The Function of different Biological Systems
4. Plants

Mathematics

1. Algebra
2. Calculus
3. Geometry
4. Trigonometry
5. Data Analysis & Statistics

Chemistry

1. Single particle systems
2. Many-particle systems (Molecular model):
3. Many-particle systems as electron pair and proton exchangers
4. Many-particle systems under thermodynamic & kinetic control

Geography

1. Physical Earth Systems & Earth-Shaping Forces
2. Human Geography
3. Globalization

English

Reading Comprehension

Themes include: family, technology, health, culture, natural phenomena, human values, environment, sports, media, arts and social issues

Writing

Writing for a variety of audiences, purposes and in various forms to communicate meanings clearly and accurately.

French

Compréhension de la lecture

Le thème : société, culture, environnement, morale, sport, médias, famille, art,...

Expression écrite : Ecrire pour un lectorat varié sur différents thèmes de manière claire et concise.

اللغة العربية

أولاً: النطاق: ماهية الأدب، المقالة، السرد القصصي، النقد الأدبيّ

ثانياً: الموضوعات: الأدب والحياة، الأدب واستشراف المستقبل، اللغة العربية وعلوم العصر، اللغة

العربية والعولمة، العولمة والانتماء، العائلة وتكريس القيم الإنسانية، القيم الإنسانية والتطور

الحضاريّ والتكنولوجيّ، الإعلام العربيّ والبعد الثقافيّ المعرفيّ

ثالثاً: النحو والصرف

رابعاً: البلاغة

خامساً: الأنماط والتقنيات

سادساً: موسيقى الخطاب (الموسيقى الداخلية والخارجية)

Figure 2: Themes covered in the pilot

A total of 3087 students enrolled in grades 10, 11 and 12 at 30 private schools in Egypt, Jordan, KSA, and Lebanon participated in the pilot. Countries where different streams are available in specific grades, only students from scientific streams participated in the pilot. After standard data cleaning, 2887 participants, 45% of which were female, were left to be included in data analysis as shown in the table below.

Table 1
Student distribution

	Egypt	Jordan	KSA	Lebanon	Total
Grade 10	150	80	270	820	1320
Grade 11	15	61	337	605	1018
Grade 12	9	32	204	304	549
Total	174	173	811	1729	2887

RESULTS

Descriptive and inferential statistical analyses were conducted using SPSS classical models and IRT Rasch models. This report is restricted to results of classical analysis.

Overall raw scores averaged below 50% on all math and science exams in all grades, as well as in French. They averaged a few points higher in English and geography. Such performance was somewhat expected because of at least three reasons. First, the IAB framework (PSE) is novel and unique in many ways, which most secondary schools around the world may not be used to. IAB drives for experiential development of a profile consisting of a balanced mix of conceptions, processes and dispositions that are critical for the development of the big picture within and across disciplines. Contrary to what educational reform has been calling for in the last two decades, including in the Arab World, instruction and assessment around the world continue to focus more on conceptual than any other aspect of covered materials, and to pay little attention, if any, to the big, especially cross-disciplinary picture. Second, all IAB exams focus more on meaningful learning rather than rote learning of course materials, with special attention to the five skill categories mentioned above. Secondary school students are not used to being tested on cognitive learning outcomes. Third, the pilot had a dual objective: to get necessary baseline data and to ascertain IAB framework and profile. In order to get a comprehensive picture of participating students' cognitive state, the level of complexity and difficulty was somewhat stretched on the eight tests towards the higher end of the cognitive spectrum. Items of various degrees of complexity and difficulty were included, with a focus more on higher-order thinking that is characteristic of meaningful, experiential learning than on routine operations and recall of materials that could be learned by rote.

Schools are provided with normalized scores of their students to facilitate data comparison with this report. Scores of all participating students on a given exam (i.e., field) in a given grade are normalized to bring the overall average on that exam to 50 points, out of 100 points, with a standard deviation of 10 points. Figure 3 shows data comparison before and after normalization in a particular exam. The reader may easily notice that normalization preserves all data properties while facilitating data comparison, especially between scores of a given student on various tests.

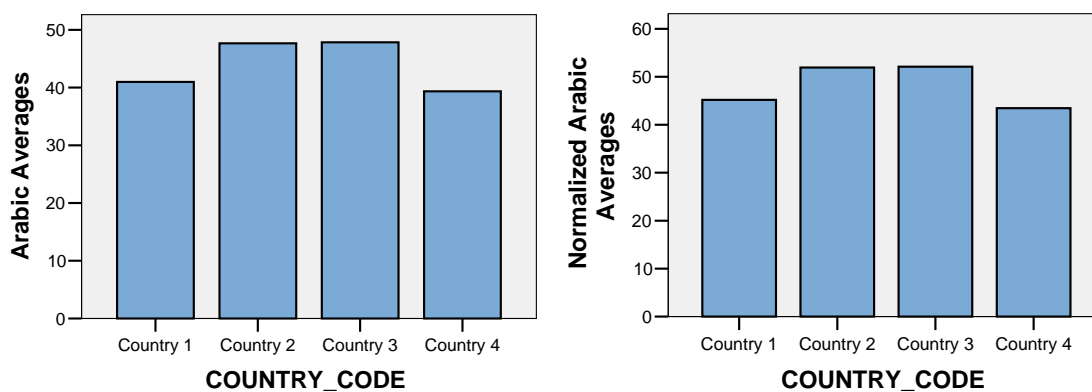


Figure 3: Country comparison of raw scores and normalized scores on a given test showing how normalization preserves data properties when it comes to statistical analysis

In the following, we outline major results of classical descriptive and inferential statistical analysis carried out with SPSS.

1. Student average performance ranked slightly different on various exams in the three grades. However, overall, and in the order shown below, student performance was best in English, and poorest in physics:

Physics < French < Chemistry < Math < Biology < Arabic < English

Furthermore, student performance on every single exam was better on epistemic items than on cognitive items.

2. We relied on simple and nested ANOVA to ascertain the effect of specific variables on student performance on various tests, and the interplay between these variables.

In statistics, ANOVA (Analysis of Variance) provides a statistical test of whether or not the averages of several groups are significantly different, when students are grouped according to a specific variable. ANOVA then tells whether or not that variable had a significant effect on student performance. The effect is quantified in terms of a coefficient F, and a probability p, the maximum value of which is 1.

For example, in Figure 3, ANOVA can help us find whether student performance on a given test is affected by the country they belong to. When such statistical test was carried out on the four groups in this figure, F turned out to be equal to 85, and $p = .000$. The extremely high coefficient and the extremely low probability indicate that in fact, in this case, performance on the given test is significantly affected by the country to which students belong. We took this case to illustrate for the inexperienced reader how we may come to certain judgments based on so-called inferential statistical tests, like ANOVA.

When simple ANOVA was carried out on individual exams within each grade, it revealed that country and school significantly affected student performance on various exams (all p-values = 0.000).

3. When *nested* ANOVA was carried out, it revealed that school and gender contributed together to the significant difference in student performance on various exams. When the influence of gender was investigated, it revealed the following within each grade:

Grade 10	Female performance > Male performance except for French Language
Grade 11	Female performance > Male performance in all fields
Grade 12	Female performance > Male Performance except for Geography

On average, females scored 2%, 9%, 4%, 4%, 11%, 6%, and 2% more than males on Arabic, biology, chemistry, math, English, geography, and physics, respectively. While in French, no difference is observed in scores between females and males.

4. To ascertain the internal coherence of various items in a given exam, we carried out a reliability test using “Cronbach Alpha”. That coefficient ranged between .83 in Grade 11 geography and .90 in Grade 12 mathematics, thus revealing a significantly high reliability of all exams.
5. To ascertain the coherence among various exams, we relied on Pearson correlation coefficient. That coefficient ranged between .67 (between geography and biology, and between chemistry and English) and .74 (between math and biology). These highly significant values ($p=.000$) support the cross-disciplinary nature of our outcomes, and thus the coherence of the profile when ascertained in the context of various disciplines. This is further supported by the same coefficient measured separately among epistemic and cognitive outcomes and between the two types of outcomes. Pearson coefficient ranged between .60 on the groups of epistemic outcomes (between biology and geography) and .75 between the groups of cognitive outcomes (biology and mathematics). The coefficient measured .64 between the groups of cognitive outcomes and the groups of conceptual outcomes on different exams (biology and mathematics), and up to .77 between the two groups of outcomes on the same exam (mathematics). The last figures are especially important since they reveal the tight relation between the epistemic taxonomy and cognitive taxonomy we have adopted.
6. To ascertain to what extent student performance on IAB exams correlate with other types of exams, we managed in one of the four participating countries to have access to participating Grade 12 students’ scores on the official state exit exam in that country. Pearson correlation between IAB exams and their state counterparts ranged between .60 and .80 in various fields. This shows that the better students are prepared in the context of IAB, the better their performance on other exams, especially national exit exams.

Another indicator in this respect was revealed when we compared overall IAB country averages to TIMSS results in science and mathematics. The country order was quite the same.

It is worth noting at this point that all results above (and more) are consistent with results obtained in the first IAB pilot in May 2009. This brings us readily to the conclusions made next.

DISCUSSION

Our pilot was conducted to answer primarily two main questions:

1. Does the taxonomy of epistemic and cognitive learning outcomes specified for IAB actually correspond to a coherent student profile? In other words, to what extent outcomes specified in various fields are actually cross-disciplinary, and how they correlate with each other?
2. To what extent IAB is viable for schools in various parts of the Arab World? Can it be readily deployed in those schools in the context of existing curricula?

Results above provide ample evidence for the first question. They basically show that the profile specified is actually a coherent profile within and across disciplines. The coherence is primarily secured by the adopted taxonomy of learning outcomes which unifies the way students are assessed in various disciplines in both content and format. It is particularly secured by the common epistemic and cognitive patterns among various fields, manifested by the unique way scope and structure are defined for various systems at the epistemic level, and the novel cognitive taxonomy that concentrates on reasoning patterns which research has shown to be shared by experts in various professions.

As for the second question, score averages on various tests fall within the viability-novelty range that we have inferred from leading evaluation research conducted on major educational projects throughout the world. Such research shows that students should never average on a given test like ours below 25%. Otherwise, what is ascertained would be beyond the potential of target students, and it would thus correspond to a program that cannot be carried out within the current state of things. Our lowest average of 33% in physics matches the viability indicator in question. Furthermore, the research in question also suggests that students should not score well above 50% on average, on a given test, for the test to be ascertaining some new ideas that are worth considering in a reformed program. Our results also match this novelty indicator to a large extent. All in all, our results give a preliminary indication that IAB is a novel viable program that can be deployed in the context of existing curricula.

Thanks to the determination and dedication of our IAB Pioneer Schools, various concerned governmental and non-governmental officials and the ERC team, we are confident that, together, we can make IAB *our students' license to the future!*