Clearing the Fog: A Comparative View of Evaluation and Assessment in the Israeli Education System

By: Naama Avidan

This paper examines the mechanisms for evaluation and assessment of student achievement in the Israeli education system, particularly as they compare to other countries. Current practices, governed by the Ministry of Education and RAMA (the National Authority for Measurement and Evaluation in Education), suffer severe failings. Recommendations are outlined for reform.

The **first chapter** gives the background, detailing concepts from the field of evaluation (longitudinal data, cross-sectional research and formative vs summative assessment), as well as the under-achievement of the Israeli education system in international exams (PISA, PIRLS, TIMSS).

The **second chapter** discusses accountability: the controversy surrounding accountability in education, accountability in comparative research, the impact of "low stakes" testing and the OECD's recommendations for accountability and evaluation.

The **third chapter** describes the evaluation and assessment mechanisms in other countries (the US, England, The Netherlands, Australia and Denmark).

The **fourth chapter** takes a deep dive into the history of Israel's policy on evaluation of the education system from 2002 to 2022, while the **fifth chapter** takes a close look at the present "Current Status" model, pointing out its flaws.

The **last chapter** lays out recommendations for a better system.

Summary

Consistent measurement of students' scholastic capabilities provides information on achievement across different population groups, and makes it possible to track trends, implement course corrections, and set measurable goals going forward. Such evaluation also gives students and their parents an indication of their individual

achievement level, allowing parents to provide support to their children when needed, and to monitor the quality of instruction they receive at school. Assessment can be divided into **formative assessment**, which takes place as part of the learning process and provides real time feedback that helps spot difficulties and tailor programs, and **summative assessment**, which measures achievement at the close of a learning period (usually via grades). Evaluation can further be conducted in two ways: (i) **cross-sectional** - providing a full picture of the system at a single point in time, allowing for comparisons between groups and the direction of resources, or (ii) **longitudinal** - continuously monitoring individual students and student groups over a span of years, enabling the formation of precise personal evaluations, the identification of development trends, and the establishment of causal relationships to track the efficiency of curricular programs.

The international review conducted for this paper shows that leading education systems around the world rely on both continuous, longitudinal assessment that tracks each student's individual progress, starting from the point of entry and continuing consistently over the course of years, and cross-sectional assessments. In recent decades, there has also been much focus on accountability – stressing the responsibility of education system actors both vertically (towards administrative bodies) and horizontally (towards parents and the public).

There is some disagreement over accountability in the context of publishing test results. Proponents argue that it is appropriate to require educational staff and administrators to set goals, meet them, and be held responsible vis-a-vis parents. In their view, public transparency is an essential tool that allows parents and the public to obtain a full picture of the state of the education system. Parents are also entitled to know their children's level of knowledge relative to expectations, in order to support their children and to align their expectations with the school. Advocates of vertical accountability likewise see value in schools' accountability to the education system and its supervisory bodies, whereby professional officials set goals and monitor educational activities in order to advance the system.

Opponents of data publication, by contrast, warn of the secondary risks it entails. They argue that the pressure generated by publication of scores may lead to excessive focus on preparing for tests, harm to the educational process, diversion of resources toward tested subjects, the creation of unfair competition, and harm to weaker populations. Tests can easily cease to function as pedagogical assessment tools and become instead public ranking instruments, degrading the school climate and damaging the broader educational process — which includes non-measurable dimensions such as values education. Therefore, in their view only "low-stakes tests", in which little or no price is paid for failure, are appropriate.

Studies indicate that accountability policy actually improves scores on various tests, and even leads at times to improvements in students' actual skills, enabling them to

reap benefits in post-secondary academic achievement and sometimes earning capacity. This effect is particularly pronounced among disadvantaged populations and in low-performing schools that made a special effort to raise their standing. Despite the drawbacks of publication — such as "cramming" material ahead of tests — it appears that a balanced form of accountability, one that neither over-publicizes data nor conceals and obscures it, contributes to the advancement of the education system. Various countries use practices that balance these considerations, in line with OECD recommendations.

Several key points for optimized formative assessment emerge from the comparative review. First, utilizing digitization in testing enables especially rapid feedback (sometimes within a single day), thereby turning evaluative tests into tools that influence pedagogic planning in real time. Second, the implementation of accountability and public transparency through digital platforms that give parents comprehensive access to information on achievement, school climate, and other data, in turn drives continuous improvement and productive dialogue with the community. Finally, striking a balance between tests that are "high-stakes" (i.e., carry important consequences) and those that are not ensures that the evaluation process reflects the students' true ability, and serves as a driving force for moving the education system forward, while avoiding the adverse effects of testing (anxiety, unfair comparisons and narrow test-oriented instruction or learning).

In Israel, the agency responsible for evaluation in the education system is "RAMA", the National Authority for Measurement and Evaluation in Education. This body operates today as an independent unit attached to the Ministry of Education, reporting its findings directly to the Minister of Education. The body was established in 2005 upon the recommendation of The Dovrat Commission, one of the most comprehensive and in-depth committees established in Israel to review the education system. Among other issues, it recommended the establishment of a national authority for evaluation in education, one that would be responsible for administering national and international exams, developing assessment tools, and conducting periodic evaluation of the education system and of individual schools, publishing its findings in an annual report. RAMA was duly established, and took on responsibility for evaluating the education system, including running the traditional *Meitzav* testing framework and managing sample-based feedback surveys for systemic evaluation in additional areas. RAMA is also responsible for Israeli students' participation in international exams and surveys: TIMSS, PIRLS, PISA, and TALIS.

An examination of RAMA's activity over its twenty years of existence, and particularly in recent years, shows that it is not only characterized by inconsistency but above all fails to fulfill the purposes for which it was established – and in some respects actually undermines them by its conduct. Over the years, the scope of testing has been drastically reduced, and the exam format has been repeatedly changed, in a way that does not allow for quality longitudinal research. The situation regarding

transparency and publication is likewise troubling: Undermining a 2012 Supreme Court ruling that required the publication of results at the school level, RAMA continues to withhold test scores, arguing that their publication harms the education system. Moreover, even the data that *is* published is of little value, as it is considerably delayed – eighteen months or more after the exams are administered. In addition, the absence of any legislation anchoring RAMA's status creates instability, and makes it difficult to manage as a professionally independent body serving the goals of the Ministry of Education. Accordingly, there is a genuine disconnect in the flow of raw data from RAMA to the Ministry of Education's senior staff (aside from processed data), and the lack of access to this information prevents decision-makers in the Ministry of Education from engaging in data-based, strategic planning.

In addition to these inherent flaws, the current testing model, "*Tmunat Matzav*" ("Current Status", in place since 2023), has made things worse: Under this model, numerical scores (0–100) have been abolished in favor of a five-tier verbal rating, and reporting is done solely in comparison to "similar schools," in such a vague way as to make it impossible to obtain information about the precise level of achievement relative to the national level.

The results of international exams, which point to a decline in the achievements of Israeli students (2023 saw the lowest TIMSS results since 2007), and to widening disparities within Israel's education system, underscore the urgent need for an effective evaluation and assessment system – one that administers exams digitally, provides longitudinal and cross-sectional data, and enables schools and the Ministry of Education to use the data for formative assessment and for improving teaching practices and the education system as a whole.

Recommendations

1. Transfer to external evaluation providers: RAMA would be one amongst other providers of assessment and evaluation services, competing to carry out the evaluation. Schools would be free to choose among different evaluation bodies, which would thus be incentivized to demonstrate efficiency and professionalism in order to earn the schools' trust. The state would continue to define the binding national standards and take on a regulatory function. Adopting a competitive system for evaluation, similar to the Dutch model, would yield several valuable advantages: (a) Innovation and technological improvement, as competition encourages providers to be better than their rivals; (b) Pedagogical suitability and relevance, as schools from different sectors can choose the measurement tools that best correspond to their specific educational vision and student population; (c) Systemic resilience, as the entire system's control room is not paralyzed (as has happened in Israel time and again) when one provider runs into difficulties or suspends its operations;

(d) Savings – as with any monopoly, opening the market is expected to lower costs (RAMA's 2025 budget stood at over 46 million shekels).

The data from the various bodies would be consolidated into a transparent digital platform (similar to the Dutch "Windows for Accountability" system), presenting the public with comprehensive school profiles. This information would allow the Ministry of Education to move to a model of differential oversight: targeted identification of low-achievement, at-risk schools, in order to provide them with close professional support and guidance, while granting greater operational latitude and reducing the bureaucratic burden on institutions that demonstrate consistent success.

2. Anchor RAMA's legal status and grant it institutional autonomy: If measurement remains under RAMA alone, it is highly important to enact a RAMA law that would enshrine its statutory status, professional independence, and powers. This would ensure stability in its measurement policy and insulate it from frequent changes, as is required of any national body responsible for measurement.

   RAMA would be obligated to provide all the data in its possession to the Ministry of Education, in contrast to the current extreme situation, in which RAMA passes on only partial, processed information. Consideration should also be given to placing RAMA under the authority of the Prime Minister's Office, thereby severing the problematic link between the evaluating body (RAMA) and the system being evaluated (the Ministry of Education).

3. Shift to consistent, transparent and formative assessment: A model of continuous assessment is needed, under which every student is tested each year in several subjects. Collecting longitudinal data of this kind would allow teaching staff to build formative assessments, set goals for each student, and track their development. Parents would receive information about their children's exam achievements individually, at least once a year, and a system of formative assessment would be in place that draws conclusions and advances each student relative to their own individual standing. To this end, the frequency of testing should be increased, and exams should be made consistent and reliable, particularly in core curriculum subjects, in line with the practice in advanced countries.

   The data would also allow for tracking broader trends – at the school, local-authority, or specific population-group level – and for assessing the impact of various intervention programs on these trends.

   To counteract the possible negative ramifications of publicizing test results, the following is recommended, modelled on the Danish and British practices: While school achievement results in grades 1–5 and 7–8 would be fully reported only to students' parents, teachers, local authorities, and the Ministry of Education's supervisory staff, the exams marking the end of elementary school and of middle school – grades 6 and 9 – would be made publicly available. Transparency in these summative exams is important for understanding schools' educational outcomes. In this way, a balance would be

struck between publishing information and heightening the "stakes" inherent in exams on the one hand, and maintaining continuous, formative measurement on the other. The education system, school staff, and parents would receive ongoing, accurate information about every student, each year, across several subject areas, while the publicly published results of graduation grades enable a comparison between schools, also highlighting underperforming schools for The Ministry of Education and local authorities.

4. Implement balanced public accountability: The achievement results of schools in grades 6 and 9 would be presented along three axes: 1) Compared to other schools nationally; 2) Compared to the school's own past achievements; and 3) Compared to similar schools. The current data does not take into account the considerable variation between student populations across different schools, thus doing an injustice to schools with students from disadvantaged backgrounds who, relative to their starting point, achieve respectable results. The exams also fail to address the full scope of a school's educational work, covering instead only certain learning domains.
   Unlike RAMA's vague definition of "similar schools", all schools in the country would be divided into 3–4 clusters based on their "nurture index" [socioeconomic index], in a fixed and transparent manner that does not conceal the reference group against which they are being compared. Each school's data would be normalized relative to its cluster of similar schools. This axis would provide a comprehensive picture that accounts for background, while still incentivizing schools to demonstrate progress.
5. Increase vertical accountability: Underperforming schools should be further monitored in order to offer them a turnaround program, whether through "*Gefen*" or other channels, ensuring that these schools do not remain at the bottom of the achievement scale. At the same time, schools that have achieved excellent results should be met with programs aimed at further enrichment and challenge suited to their students' high abilities.
6. Increase incentives for students' participation: As the research demonstrates, the absence of "stakes" for students leads to a loss of motivation, and as a RAMA representative herself noted, there is an opposite incentive for students to minimize effort, finish the exam early and go out to recess. The practice of anonymous testing is not internationally accepted, and is not used for other exams that students take at school. Deliberately increasing the "stakes" of evaluative exams for students, thus increasing their motivation to participate and succeed, is actually beneficial, and can be achieved by reporting the scores to parents, and giving the students feedback in their report card: a verbal report in grades 1–4 and a numerical score in grade 5 and above.
   Concerns about manipulation or the exclusion of weaker students from testing should be addressed in the same way they are addressed with respect to any other exam. Furthermore, the more common government exams become, the more the tension surrounding them will diminish. In many countries around the world, national exams are held regularly across several subject areas every

year — sometimes even several times a year — making them a routine matter and a natural part of the learning process, without generating the pressure, or driving students to external prep centers, that some claim.

7. Complete digitization and rapid feedback: The transition to fully online evaluation exams using artificial intelligence tools should be completed as quickly as possible. The use of technology would allow for a dramatic reduction in the time it takes to receive feedback. Such digital exams, graded with the aid of artificial intelligence, already exist today within RAMA itself. These exams are called "*Tnufa* diagnostic tests" and they allow teachers to conduct assessments of their students. These exams should become mandatory in every class, in each of the core curriculum subjects (with additional exams developed as needed), and their administration overseen by professional bodies, with the resultant data channeled to both parents and education personnel.
8. Utilize novel evaluation and assessment tools: In an era when digital learning software is used for the learning process itself, and the Ministry of Education is moving toward implementing the "720" initiative that will collect up-to-date information on each student, efforts should be made to synchronize the data from those technology platforms as they are approved and operated, in cooperation with the Ministry of Education and the Ministry's data repository. The more comprehensive and up-to-date this information is, the better the Ministry of Education will be able to formulate strategies and make the most of students' potential. Needless to say, this information must be collected without compromising students' privacy.

See the full Hebrew paper

(See also "The Israeli Education System's Broken Compass: Evaluation and Assessment in the Education System")