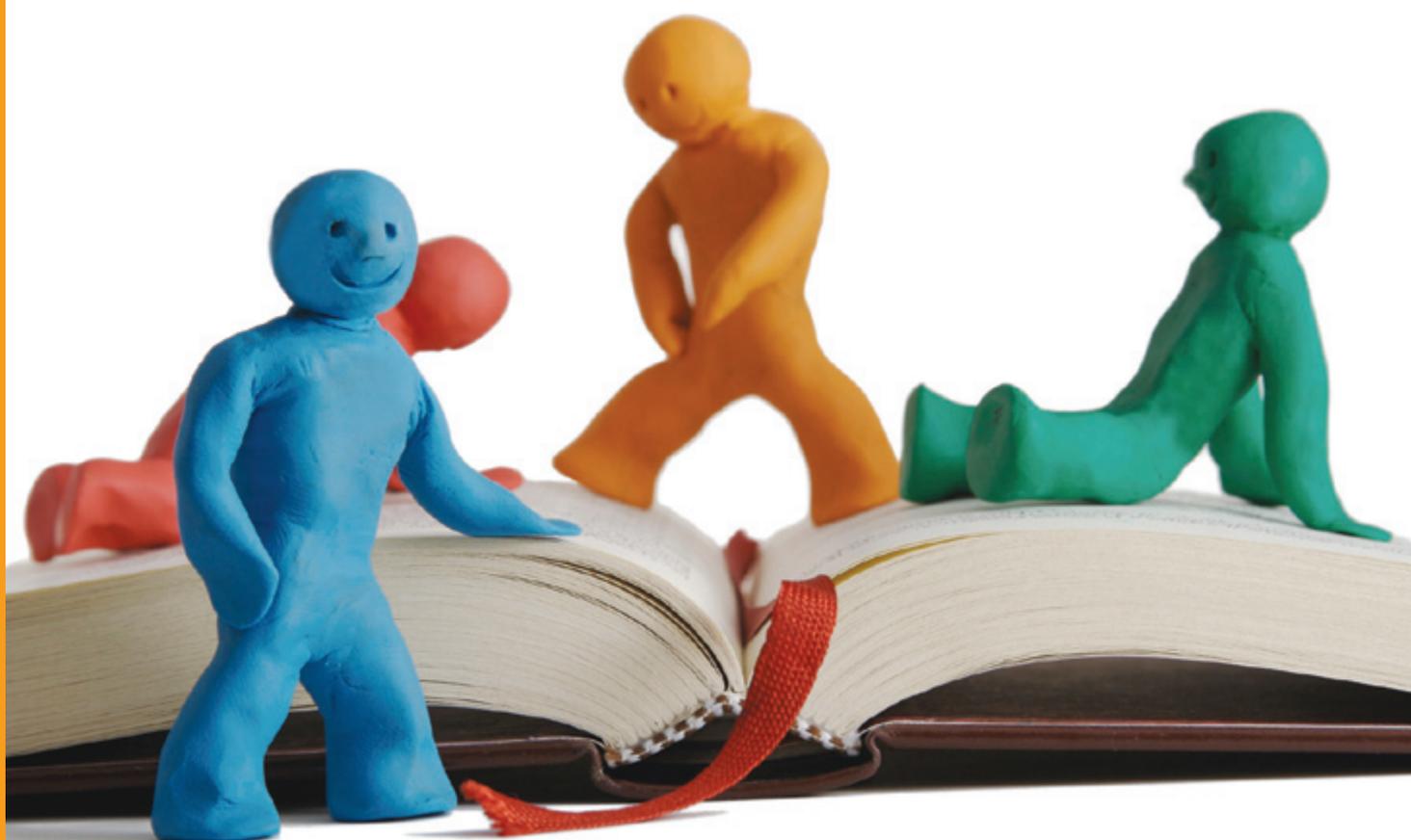


Diagnosis, Assessment and Evaluation in Preschool through Second Grade

Status Report and Recommendations
by The Investigative Panel on Diagnosis, Assessment
and Evaluation in Early Childhood Education

Edited by Zvia Breznitz and Reut Yamin



Applied Research in Education

Diagnosis, Assessment, and Evaluation from Preschool to Second Grade

Status Report and Recommendations

by the Review Panel on

Diagnosis, Assessment, and Evaluation in Early Childhood Education

Initiative for Applied Research in Education

Israel Academy of Sciences and Humanities

Ministry of Education

Rothschild Foundation (Yad Hanadiv)

Diagnosis, Assessment, and Evaluation from Preschool to Second Grade

Status Report and Recommendations

by the Review Panel on

Diagnosis, Assessment, and Evaluation in Early Childhood Education

Edited by

Zvia Breznitz and Reut Yamin

Jerusalem, 2008

Initiative for Applied Research in Education

Israel Academy of Sciences and Humanities

Academic editor and copyeditor of the Hebrew edition: Kari Druck
Israel Academy of Sciences and Humanities
www.academy.ac.il

ISBN 978-965-208-193-3

© 2008 by Israel Academy of Sciences and Humanities. All rights reserved.

A few of the diagnostic tools listed in the appendix are not included in the reference list. Despite intensive efforts, we were unable to locate these references. If we find them in the future, we will add the information to the next edition.

Cover design: Shimon Schneider Studio

The Israel Academy of Sciences and Humanities was founded in 1959. Its membership currently comprises ninety-four top Israeli scientists and scholars. According to the Israel Academy of Sciences and Humanities Law, 1961, its principal objectives are to foster and promote scientific activity; to advise the Government on research activities and scientific planning of national importance; to maintain ties with equivalent bodies abroad; to represent the Israeli scientific world at international institutes and conferences; and to publish articles that can further scholarship.

The Ministry of Education was founded in 1948, when Israel became an independent, sovereign state. Under the State Education Law, 1953, the Ministry is responsible for the education of Israeli children from preschool through high school, up to and including the twelfth grade. In addition, the Ministry is in charge of teacher-training in colleges of education. The Ministry sets both pedagogical policy (e.g., development of curricula, teaching methods, and standards) and organizational policy (e.g., budgeting for the education system, logistical planning, attention to special population groups, and inspection of educational institutions).

The Rothschild Foundation (Yad Hanadiv) is continuing the Rothschild family's philanthropic activity in Israel, which Baron Edmond de Rothschild began in the late nineteenth century. The Rothschild Foundation works on improving educational achievement, especially by increasing opportunities for all Israeli pupils to receive a high-quality education. The Rothschild Foundation makes cutting-edge knowledge and expertise available to education workers, thereby spurring innovation, which can improve vital components of the Israeli education system.

The Initiative for Applied Research in Education was founded in late 2003 as a joint project of the Israel Academy of Sciences and Humanities, the Ministry of Education, and the Rothschild Foundation, in order to develop applied research on issues of concern to decision-makers in the field of education. Its aim is to help decision-makers improve educational achievement by providing them with the latest peer-reviewed knowledge. The Initiative is based on the experience of the United States and Europe, where national academies have undertaken to advance the education systems by learning from both research and practice. In these regions, an improvement in student achievements has been related, under certain conditions, to the systematic use by teachers, principals, and policymakers of knowledge and evidence derived from scientific research.

Three working assumptions guided the establishment of the Initiative:

- New knowledge in various fields of research, from brain science to performance, may contribute to research and practice in education. In Israel there are research capabilities—in education and other fields—that can be encouraged to focus on improving educational achievement.

- Asking research questions derived from decision-makers' agenda may encourage education researchers, in collaboration with scholars in other fields, to expand their activity aimed at creating knowledge that can benefit education practitioners. The effort to answer these questions may lead to new theories and research tools for advancing the education system and education.
- Decision-makers in education, from teachers to the Education Ministry administration, will seek to derive benefit from carefully reviewed and established knowledge that is made available to them and to contribute to the development of such knowledge based on their own professional experience.

The Investigating Panel on Diagnosis, Assessment, and Evaluation in Early Childhood Education

The Investigating Panel on Diagnosis, Assessment, and Evaluation in Early Childhood Education was formed in response to a request by the director general of the National Authority for Measurement and Evaluation in Education (RAMA), which is in charge of policymaking regarding assessment and evaluation in the Israeli education system (I prefer educational system, education is a noun, not an adjective). The director general of RAMA and the Education Ministry administration asked the Initiative for Applied Education Research to survey the present situation in Israel in these respects and to provide them with the best practical and research-based knowledge to help them formulate effective policy for early childhood education. Early childhood is generally defined internationally as lasting from preschool age through the first three grades of elementary school (ages 3–9). However, the Investigating Panel chose to include in its purview preschool through second grade (ages 3–8) only, mainly because Hebrew-speaking schools in Israel make the switch from pointed Hebrew to unpointed Hebrew after second grade. This switch requires mastery of advanced language and linguistic literacy skills, and the acquisition of these skills has to be examined separately.

The work of the Investigating Panel follows up the activity of two previous committees formed by the Initiative for Applied Education Research: the Committee on Assessment and Evaluation in Education and the Committee to Review Modes of Education in Early Childhood. The Committee on Assessment and Evaluation in Education summed up its work in a report entitled “The Knowledge Base for Assessment and Evaluation in Education: A Framework for Curricula” (2005). In this report the committee defined basic terms relating to diagnosis, assessment, and evaluation; emphasized the importance of evaluation as a vital component in the teaching-learning-evaluation cycle; and discussed the knowledge needed by diagnosticians, assessors, and evaluators to carry out their educational work and/or research. The Committee to Review Modes of Education in Early Childhood summed up its work in a report entitled “From Research to Practice in Early Childhood Education” (Klein and Yablon, 2007). In the report, the committee advocated promoting the following aspects of child development in early childhood: (1) social and emotional adjustment; (2) language, linguistic literacy, and mathematical skills; (3) openness to learning and creativity. The committee recommended monitoring the development of children in these areas, partly in order to understand the contribution of the education system to children’s development. Such

understanding requires evaluation of their achievements, as well as of the quality of the interaction between teachers and children, a variable that many studies have found to be the best predictor of children's future achievements. The committee also recommended including relations between parents and children and between parents and educational institutions in the diagnosis.

The Investigating Panel focused on diagnosis, assessment, and evaluation in early childhood of social and emotional development and of the acquisition of language, linguistic literacy, and mathematical skills. Initial inquiries with senior officials in the education system, in the Psychological Counseling Service, and in the field of education indicated that various organizations, both governmental and private, are active in these areas, and that each of them employs its own methods to further its particular goals. In most cases the diagnostic, assessment, and evaluation tools used lack countrywide norms, and it is therefore questionable whether reliable, valid information can be obtained from them. Consequently, a survey of the present situation in the education system was needed. In this survey, the panel identified the needs of the education system with respect to diagnosis, assessment, and evaluation in early childhood (3–8), carried out a comprehensive survey of the processes and tools used by the education system and of the abilities accessible to it, identified deficits, and recommended ways of addressing the deficits.

The present report sums up the work of the Investigating Panel in accordance with the purposes for which it was formed. The structure of the report follows the work process: first, the needs of the Education Ministry and RAMA were studied and the processes and tools currently used by the education system for diagnosis, assessment, and evaluation were surveyed. Next, the panel described various aspects of the present disparity between real and ideal in diagnosis, assessment, and evaluation of socio-emotional adjustment and language, linguistic literacy, and mathematical skills. Afterwards the panel made specific recommendations for each of these fields, general recommendations for further research and development in these fields, and recommendations in principle regarding standards for development, implementation, distribution, and introduction of appropriate tools for diagnosis, assessment, and evaluation in the education system.

Our hope is that this report and its recommendations will help those involved in the field to become familiar with the present state of affairs regarding diagnosis, assessment, and evaluation of children in preschool through second grade and will serve as a ready source of knowledge that can be used for effective, professional, comprehensive diagnosis, assessment, and evaluation in the education system.

Members of the Investigating Panel:

Prof. Zvia Breznitz, University of Haifa (chair)

Dr. Anat Ben-Simon, National Institute for Testing and Evaluation

Prof. Yoel Elizur, Hebrew University of Jerusalem

Dr. Haggai Kupermintz , University of Haifa

Prof. Iris Levin, Tel Aviv University

Dr. Fadia Nasser Abu al-Hija, Tel Aviv University

Prof. Dorit Ravid, Tel Aviv University

Prof. Anat Scher, University of Haifa

Prof. David Share, University of Haifa

Prof. Dina Tirosh, Tel Aviv University

Dr. Reut Yamin, panel coordinator, Initiative for Applied Research in Education

Acknowledgments

The members of the Investigating Panel would like to thank the following researchers and practitioners who helped them gather data, provided information, broadened their knowledge, and offered good advice.

We thank Ministry of Education representatives who took part in the needs-survey meeting (in alphabetical order): Leah Achi-Miriam, director of the Educational Involvement Division in the Department for Elementary School Education; Avi Bar-Kochba, deputy director of the Department of Education and Welfare Services; Prof. Michal Beller, director general of the National Authority for Measurement and Evaluation in Education (RAMA); Etti Bookspan, senior instructor in the Educational Involvement Division of the Department for Elementary School Education; Dr. Yehudit Eldor, director of the Learning Disabilities Unit in the Psychological Counseling Service; Ora Goldhirsch, psychologist and national instructor in the Department for Pre-School Education; Abdullah Khatib, director of the Department for Arab Education; Rawda Kraini, inspector of Arabic language education in the Department for Curricula; Tamar Pitkhiya, the official in charge of social services and rehabilitation in the Department of Education and Welfare Services; Shlomit Rahmel, director of the Department for Talented and Gifted Students; Dr. Yoel Rapp, manager of assessment development in RAMA; Ada Rosenberg, assistant chair of the Pedagogical Secretariat; Nava Segen, director of the Department for Curricula; Maya Sharir, national instructor, Immigrant Absorption Unit; Dalia Tal, national inspector in charge of pedagogical education, Learning Disabilities Unit; Monica Winokur, national coordinator of the Ma'agan program and national instructor in the Department for Pre-School Education; Dr. Zofia Yoed, director of the Subject Matter Unit and deputy director of the Department for Curricula. The panel would also like to thank Leah Rosenberg, head of the Pedagogical Administration, for sharing her insight and her time; and Chani Shilton, head of the Formative Assessment Section in RAMA, for her assistance.

Education Ministry representatives who offered advice on socio-emotional development: Dr. Tamar Erez, Psychological Counseling Service; Dr. Sharona Meital, Psychological Counseling Service.

Consultants on mathematics: Prof. Abraham Arcavi, Weizmann Institute of Science; Dr. Miriam Ben-Yehuda, Beit Berl College, expert on diagnosis of learning disabilities in mathematics; Dr. Zvia Markovitz, Oranim College, chair of the Core Program for Preschool Mathematics (currently being developed); Prof. Pessia Tsamir, Tel Aviv University.

Education Ministry representatives specializing in mathematics: Tammy Giron, national instructor for mathematics inspection; Dr. Chana Perl, chief inspector for mathematics; Dr. Anat Sela and Sarita Spokoiny, Science and Technology Administration and Department for Pre-School Education.

Authors of background papers for the work of the Investigating Panel: Dr. Meli (Elimelech) Nafcha; Dr. Ronit Roth-Hanania, developmental psychologist.

Special thanks go to Prof. Michael J. Feuer, director of the Division on Behavioral and Social Sciences and Education at the National Research Council, for sharing his insights, offering constructive advice based on his experience, and steering the Investigating Panel in a productive direction.

We are grateful to the staff of the Israel Academy of Sciences and Humanities for all their help and goodwill in publishing the report, and to the people at the Rothschild Foundation for their marvelous, friendly cooperation and constructive criticism.

We thank Kari Druck, the academic editor and copyeditor of the Hebrew edition, for her critical reading of the report and her professional and constructive judgment.

Warm thanks to Riki Fishel and Etti Amit of the Initiative for Applied Research in Education for their professional, patient assistance to the Investigating Panel.

Peer Review

Due to the nature of the work of the Investigating Panel, a preliminary draft of the report was sent to the people who commissioned it—the director general of RAMA and officials in the Ministry of Education. The panel took their comments into account and revised the report accordingly. The report summing up the work of the Investigating Panel was sent for peer review to experts in Israel and abroad as a means of subjecting it to a critical and objective external review so as to produce the best document possible. To ensure objectivity, the reviewers' identity was withheld from the panel members until publication of the report.

We would like to thank the reviewers for their assistance:

Dr. Dorit Aram, Tel Aviv University

Dr. Hagit Benziman (retired), Hebrew University of Jerusalem

Prof. Sorel Cahan, Hebrew University of Jerusalem

Prof. Kenji Hakuta, Stanford University

Prof. Baruch Nevo, University of Haifa

Prof. Amiram Raviv, Tel Aviv University

The reviewers listed here offered constructive comments and suggested additions and revisions to the draft that they read. However, they were not asked to endorse the conclusions and recommendations of the Investigating Panel and did not even see the final version of the report before publication. Responsibility for the content of this report rests entirely with the Investigating Panel.

Zvia Breznitz, Chair

Reut Yamin, Coordinator

After having completed its work, the Investigating Panel was informed that RAMA recently developed a test of Arabic language acquisition for Arabic-speakers. The panel welcomes this development and hopes that it will be one of several improvements that RAMA and the Ministry of Education make, based in part on the panel's recommendations.

Contents

Summary of Findings and Conclusions	12
1. <i>Procedure</i>	12
2. <i>Conclusions and Recommendations</i>	13
3. <i>Specific Recommendations Regarding Functioning and Content</i>	18
Introduction	21
1. <i>Background: Formation of the Investigating Panel</i>	21
2. <i>Goals and Procedure</i>	21
3. <i>Major Milestones in the Work Process</i>	24
4. <i>Frame of Reference and Professional Considerations</i>	25
5. <i>Survey of the Needs of the Ministry of Education</i>	26
Chapter 1. Survey of Needs of the Ministry of Education with Respect to Diagnosis, Assessment, and Evaluation in Preschool through Second Grade	28
1. <i>Needs of the Psychological Counseling Service</i>	28
2. <i>Needs of the Department of Education and Welfare Services</i>	30
3. <i>Needs of the Department for Pre-School Education</i>	30
4. <i>Needs of the Department for Elementary School Education</i>	31
5. <i>Needs of the Department for Arab Education</i>	32
6. <i>Needs of the Department for Talented and Gifted Students</i>	33
7. <i>Needs of the Department for Special Education</i>	34
8. <i>Needs of the Immigrant Absorption Unit</i>	35
9. <i>Conclusion</i>	37
Chapter 2. Real versus Ideal: Implications of Evaluation Methods and Specific Recommendations	38
1. <i>Socio-Emotional Development</i>	38
2. <i>Language and Linguistic Literacy</i>	42
3. <i>Mathematics</i>	46
Chapter 3. Summary and Recommendations Regarding the Issues Examined	56
1. <i>Recommendations for Further Research and Development</i>	56
2. <i>Guiding Principles for Screening and Diagnosis</i>	60
3. <i>Fundamental Methodological Recommendations</i>	62
4. <i>Standards for Distributing and Introducing Tests</i>	66
5. <i>Diagnosis, Assessment, and Evaluation in a Multicultural Society</i>	67
6. <i>Systemic Evaluation</i>	70
References	72
<i>Appendix B: Diagnostic, Assessment, and Evaluation Tools Used in Early Childhood Education: Proposal for an Evolving “Map”</i>	76

Summary of Findings and Conclusions

The Investigating Panel on Diagnosis, Assessment, and Evaluation in Early Childhood Education was formed by the Initiative for Applied Education Research in response to a request by the director general of the National Authority for Measurement and Evaluation in Education (RAMA) and the Education Ministry administration to survey the present situation in Israel in preschools and grades 1–3 and to provide them with the best practical and research-based knowledge so as to help them optimize policy for early childhood education.

The aims of the Investigating Panel were as follows:

1. To examine diagnosis, assessment, and evaluation in early childhood in Israel from the theoretical and practical standpoints, focusing on three main aspects of development: (a) socio-emotional adjustment; (b) language and linguistic literacy skills; (c) mathematical skills
2. To survey the needs of the education system in terms of screening, diagnosis, assessment, and evaluation of typical children in preschool through second grade (ages 3–8)
3. To survey the normative evaluation tools used today in the education system and to describe the purpose of each tool, the information it provides (screening, diagnosis, evaluation of functioning), the age group for which it is intended, the way it is used, the users, the training needed to use it, and the recipients of the information obtained
4. To identify the components of the disparity between real and ideal in the education system, i.e., to identify needs that are being met and needs that are not being met
5. To recommend ways of broadening knowledge, developing appropriate tools, and improving teaching, diagnosis, assessment, and evaluation

The Investigating Panel focused on the evaluation of individuals within the education system, but it also discussed means of evaluating the system as a whole and special populations in a multicultural society. Although one aspect of a systemic evaluation pertains to teacher-student relations, the Investigating Panel did not focus on this issue.

1. Procedure

The Investigating Panel started working in November 2006. Its work comprised the following stages:

1. The panel met with representatives of the various Education Ministry departments to survey the needs of the education system and discover what exists and is missing.
2. The panel looked at various tools and processes currently available to the education system for diagnosing, assessing, and evaluating children in preschool through second grade. It examined

their psychometric validity and reliability, the existence of norms, and the ability of the tools and processes to meet needs with respect to diagnosing, assessing, and evaluating the functioning of individuals and the system.

3. Afterwards, the panel drew conclusions and formulated recommendations.

2. Conclusions and Recommendations

The panel's overall impression based on its review of the present situation was that in most respects there is a shortage of normative screening, assessment, and evaluation tools in Israel that would make possible intelligent, systematic examination of early childhood development on the level of the individual and on the level of the system. The situation is especially dire in the Arab sector.

The panel relied on two basic assumptions:

1. Evaluation is systematic observation of children's abilities, functioning, and achievements for the purpose of counseling, progress, and improvement. Any evaluation must be based on the educational goals whose achievement it is meant to examine. Consequently, the Investigating Panel treated evaluation not as an end but as a means to helping children progress in various ways and various senses.
2. Evaluation should focus both on the individual, for the purpose of identifying, surveying, and diagnosing his/her unique difficulties and abilities, and on the system, in order to draw optimal conclusions regarding the educational environment, teaching methods, the level of instruction, the quality of teaching, and the impact of these factors on the individual's learning.

2.1 General Conclusions

1. An evaluation of the implementation of new curriculum units should be an integral part of every educational initiative. The education system should develop systematic, orderly research mechanisms for monitoring changes over time in preschools, classes, and schools following educational interventions such as the introduction of new curriculum units. Every educational intervention should have a formative evaluation, and only after the intervention has become entrenched and stabilized should a summative evaluation be done. Such a course of action can enhance the stability of the system and lead to changes based on genuine needs and on well-defined goals and expectations.
2. There is a disparity between the development of research-based knowledge and the application of this knowledge in the development of diagnostic, assessment, and evaluation tools. Tests currently used by the education system do not take advantage of a large quantity of research-based knowledge that has accumulated. Mabatim, a tool developed at the national level and

currently being disseminated in the education system, is now being used for educational/developmental evaluations in many Israeli preschools, both Hebrew-speaking and Arabic-speaking (a version of the tool was developed specifically for Arabic-speakers), based on observations by the preschool teacher. There are salient advantages to this tool (for details, see chapter 3). However, researchers and academics were not involved in its development, and this fact is noticeable in the deficient description of abilities expected of the children in all areas of development. In addition, the lack of norms greatly lessens the ability to draw conclusions regarding children's abilities and to make decisions regarding their functioning.

3. The education system currently uses many tools that are based on unclear rationales, do not meet standards of validity and reliability, and lack norms. The Investigating Panel focused mainly on valid, reliable tools with a clear rationale and countrywide norms.
4. There is a noticeable gap among teachers between knowledge of theory and adoption of steps for implementation. The shortage of normative tools makes it difficult for educators in preschools and schools to connect theory (knowledge about child development) and practice (evaluation of a particular behavior in a particular child in order to determine whether it is typical or exceptional). Development of normative tools should help teachers (a) understand what is a normal level of functioning and what is exceptional behavior; (b) identify students whose achievements are exceptional (gifted students and students having special difficulties); (c) assess the quality and impact of intervention programs; (d) monitor changes over time.
5. As the next step following the accumulation of extensive research-based knowledge about early childhood development, a comprehensive, systematic study is required that directly charts the development of children's various abilities during this period (especially in mathematics). Such a study is, of course, essential for the development of diagnostic, assessment, and evaluation tools as well.

2.2 General Recommendations

1. There should be an effective agency in charge of brokering and mediating between researchers and the education system. This agency would pass on research-based knowledge to the people in charge of curriculum development and curriculum evaluation in the Ministry of Education.
2. A professional agency should work with the age divisions (preschool, elementary school, etc.) in the Ministry of Education to coordinate diagnosis, assessment, and evaluation of all aspects of education. This agency would channel collaboration, help prevent redundancy, pool resources, and promote two main focal points: (a) development of diagnostic, assessment, and

evaluation tools; (b) development of normative indices in accordance with information obtained through research.

3. This professional agency should appoint professional subcommittees as needed for the various fields of content. Each subcommittee should include experts in the field of content and in methodology. These subcommittees would be subordinate to the agency and would be run by it.
4. Israel already has a central agency (RAMA) responsible for coordinating assessment and evaluation in the education system. It is worth considering the possibility of having it develop tools for practical educational evaluation in a multicultural society and among special-needs populations and oversee their implementation.
5. The tools developed should be for purposes of identification, screening, diagnosis, assessment, and evaluation of various aspects of functioning, including socio-emotional adjustment and, to a certain extent, mastery of language skills, linguistic literacy, and mathematics.
6. These tools should be geared toward:
 - a. The level of the individual, in an effort to address difficulties and evaluate achievements—examining children’s abilities and achievements and identifying and diagnosing children at risk, children who are having difficulty or failing, and gifted and talented children. Depending on the structure of the tool, it may be used in a group or class setting or with individuals.
 - b. The level of the system, in an effort to evaluate teaching and education and the factors that determine their quality. The tools should look at preschool teachers, schoolteachers, curricula, diagnostic indices, achievements in preschools and schools, achievements in different localities, and systemic development processes.
7. The tools should be developed for diagnosis, assessment, and evaluation of various population groups in a multicultural society. The panel emphasizes that it found a *total absence* in the Israeli education system of tools with norms for screening, diagnosis, assessment, and evaluation in the Arab, Druze, and Bedouin populations. The panel recommends urgently rectifying this regrettable state of affairs. A clear, precise definition of the aims of any diagnosis, assessment, or evaluation geared toward treatment and intervention and designation of the identity of the examiners must precede any diagnostic, assessment, or evaluation process.
8. The panel notes that the evaluation policy in a multicultural context must focus on developing unique, independent assessment tools designed to meet the educational and cultural needs of

each sector of the population. The assessment tools should not rely on translating similar tools from other languages and cultures.

9. The tools should be developed in accordance with accepted methodological principles for the development of tests and evaluation tools. Their validity, reliability, fairness, and use in education should be examined by means of research.
10. Development of the entire field of diagnosis, assessment, and evaluation in education must be based on clear, well-defined rules of ethics, backed by a legal system. The legal system, along with educational agencies, should also discuss whether parental permission should be required for diagnosing children for the purpose of treatment. In addition, procedures should be established for parental involvement in the diagnostic process, the drawing of conclusions, and the planning of methods of intervention.
11. During the development of a screening, diagnostic, assessment, and evaluation system, the target population should be divided into four subgroups: pre-nursery (ages 3–4), nursery school (ages 4–5), compulsory preschool (ages 5–6), and grades 1–2 (ages 6–8). The tools for testing children in these subgroups must be development-appropriate.
12. Diagnosis of learning readiness in preschools should be examined in depth. At present, readiness is determined mainly by psychologists from the Educational Psychology Service, based on a variety of tools (e.g., Bender, Wechsler, House-Tree-Person, CAT). The quality of decision-making in this field should be examined through in-depth studies providing findings to guide the development of indices of school learning readiness.
13. Developing tools for identifying children at risk from age 3 may spare the children and their families suffering and may save resources of the Ministry of Education.
14. For each tool developed, the following should be decided in precise terms in advance:
 - a. Objectives and goals: Is it meant to evaluate functioning? To test for difficulties and/or failure to acquire cognitive or socio-emotional skills? To evaluate achievement in a particular subject for the sake of treatment and planning of interventions?
 - b. What examiners/evaluators are entitled to use the tool and what training should they have?
 - c. How should the tool be administered in practice? What are the procedures for analyzing, interpreting, and drawing conclusions from the data?
 - d. To whom should the information obtained from use of the evaluation tool be provided (another education system, people within the education system in which the child is enrolled, the family, doctors, allied health professionals, etc.)?

These questions should be answered in the wake of consultation with the Education Ministry departments (depending on what unit deals with the content of each tool) and with the child's parents.

15. A monitored system should be set up to train users of the various tools. Permission to use a tool should be based on need. Training should focus on (a) the importance of screening, diagnosis, evaluation, and assessment in early childhood; and (b) familiarity with the tool, including the logic behind it, use of it, and its effects. The training should be for teachers in preschools and the lower grades of elementary school, for guidance counselors, and for other professionals who work in early childhood education.
16. In order to carry out evaluations in practice, each preschool or school should have a coordinating body whose job it is to coordinate all diagnostic information gathered about individual children, individual staff members, and the system as a whole. The coordinating body should ensure that changes are made, educational interventions are implemented, and treatment is provided as needed. The Investigating Panel recommends that the preschool teacher, homeroom teacher, or school learning evaluators assess individual children, with assistance from psychologists and guidance counselors. The principal or inspector of preschools should carry out systemic evaluations of the school or preschool.
17. It is important to create a database of the information gathered in the various fields of content. This database should be available to researchers and developers so that the education system can become more productive. The Investigating Panel recommends that this database be the responsibility of RAMA.
18. The Investigating Panel was instructed to describe the present situation with respect to diagnosis, assessment, and evaluation in early childhood, but not to recommend priorities for developing diagnostic, evaluation, and assessment tools. The panel suggests that the scale of priorities for developing the tools that are lacking be determined by the authority that is in charge of all aspects of diagnosis, assessment, and evaluation in early childhood and recommends that RAMA be this authority.
19. In view of the numerous realms in which diagnosis and evaluation are required at a young age, there is a risk of excessive evaluation. Excessive evaluation may be a burden on the educational staff, which has to put time and resources into evaluating instead of teaching, may interfere with the students' regular activities, may lead to intensive preparation of students for the diagnosis/evaluation, and may cause tension among students and teachers alike. Therefore, the developers of the model of diagnosis/evaluation should consider distributing activities over time, coming up with methods that are not demanding in terms of time, and if possible the

inclusion of activities that are similar to the regular activities carried out in preschool or in class.

3. Specific Recommendations Regarding Functioning and Content¹

The Investigating Panel's recommendations only addressed aspects of development in which the panel members have expertise. The panel did not look at other aspects of development (e.g., motor development, curiosity, creativity, and problem-solving ability) and other fields of content (e.g., music, art, and science and technology) that are no less important; these are worth exploring in the future. In this context, it is important to note that the situation in Israel in terms of the evaluation of young children's intelligence is unsatisfactory. The IQ tests currently used by placement committees to decide children's fate use outdated norms, and up-to-date standardization is needed. The situation is especially problematic with respect to Arab, Druze, and Bedouin children and haredi children.

3.1. Socio-Emotional Development²

It is important to note that we are lagging far behind the West in terms of evaluation of socio-emotional aspects of behavior. One manifestation of this is that psychologists rely on invalid tests that require of them an unreasonably large investment of time. The panel recommends adopting and developing standardized tools that make it possible to gather information from parents and teachers as part of an orderly diagnostic and decision-making procedure for children in the education system.

3.1.1 Individuals

- a. Tools should be developed with multicultural Israeli norms that meet criteria of validity and reliability in order to help psychologists and guidance counselors in the Psychological Counseling Service streamline the identification of children having special difficulties.
- b. The diagnostic tools should be developed in a similar fashion. We suggest focusing at present on two tools—one relating to socio-emotional behavioral problems such as emotional regulation and social functioning, and the other specifically diagnosing ADHD.
- c. Diagnoses should be made by specially trained educational psychologists and should include structured, efficient gathering of information from parents and educators. The information should be gathered with the consent of the child's family and the conclusions of the diagnosis

¹ For details, see chapter 2.

² For details, see the section on socio-emotional development in chapter 2.

should be conveyed to the parents. Only with the parents' consent should the information be provided to other educational personnel.

3.1.2 *The System*

- d. There is a need for the development and implementation of tools and observation methods for evaluating the educational environment and its suitability to the children's developmental and emotional needs. Tools of this sort already exist and can be adapted for use in Israel.
- e. Based on the Klein committee report "From Research to Practice in Early Childhood Education" (Klein and Yablon, 2007), which showed, by means of research findings, that interactions between teachers and students are the best predictor of improvements in students' achievements, the panel suggests promoting the use of valid, reliable tools for evaluating these interactions (see, e.g., Pianta, 1992, 1995).
- f. An evaluation tool should be developed that focuses on the interrelations between the child, the family, the education system, and health and social services in such a way as to make it possible to assess what is available and what is lacking in this regard.

3.2 *Language and Linguistic Literacy*³

3.2.1 *Individuals*

- a. There should be a focus on developing tools to assess skills pertaining to the spoken and written language, reading, writing, and text composition for the four age groups: pre-nursery (ages 3–4), nursery school (ages 4–5), compulsory preschool (ages 5–6), and grades 1–2 (ages 6–8). The diagnostic tools should be developed in keeping with the language that the child uses in preschool/school (Hebrew or Arabic) and should take into consideration the child's mother tongue, which may not be the language of the preschool/school. In addition, the diagnostic tools should be adapted to the curriculum.
- b. The initial diagnosis should be made by the teacher by means of screening tests for all pupils in an effort to identify those having difficulty and evaluate achievements so as to be able to teach effectively.
- c. The second stage is a professional diagnosis, which should be carried out by the learning evaluator or other experts, depending on the type of problem identified.
- d. Tests in the preschool years should be administered individually or in small groups. In elementary school, they should be administered to entire classes or individually as needed.

³ For details, see the section on language and linguistic literacy in chapter 2.

3.2.2 *The System:*

- e. Tools (questionnaires and tests) should be developed for evaluating the children's literacy environment, instruction, and the relevant learning outcomes. Also worth noting is the importance of a follow-up study on the implementation of the language and literacy curriculum.

3.3 *Mathematics*

3.3.1 *Individuals*

- a. Given that there are no suitable diagnostic indicators of mathematical skills in early childhood, a comprehensive study should be undertaken to obtain information on normative mathematical functioning in terms of knowledge, subject matter, and mastery of skills among children in preschool through second grade.
- b. Based on the information obtained, the following should be developed:
 - 1. Tests (age- and content-appropriate) with multicultural Israeli norms that meet criteria for validity and reliability and measure interim and final outcomes in terms of mathematical knowledge and thinking processes of children in preschool through second grade. The tests should also examine children's attitudes toward mathematics, because positive attitudes have been found to prevent emotional and academic difficulties later.
 - 2. A standardized diagnostic tool for evaluating quantitative cognitive functioning that can be used to identify a mathematical disability (dyscalculia).
 - 3. A tool for identifying children who are gifted in mathematics.
- c. These tests should be administered by the preschool or elementary-school teacher to individuals, small groups, or the class as a whole.

3.3.2 *The System*

Tools (questionnaires and tests) should be developed for evaluating teaching and the outcomes of studying math. Also worth noting is the importance of a follow-up study on the implementation of the math curriculum.

Introduction

1. Background: Formation of the Investigating Panel

Implementation of the 1984 amendment to the Compulsory Education Law extending free, compulsory education to all children aged 3–4 has made it necessary for the Ministry of Education to engage in systematic monitoring of educational activity in preschools, and to this end to diagnose, assess, and evaluate the functioning of young children. Although numerous studies have found that early childhood education has an impact on learning in school, there is no consensus regarding the ability to assess children in this age group properly. Some people maintain that the tremendous diversity in the pace of child development makes it impossible to set norms for evaluating development or even to predict future levels of functioning. On the other hand, it has been found that evaluating emotional regulation and effortful control at a young age, identifying behavioral problems and evaluating their severity, and keeping track of mediating variables such as IQ and educational achievement make it possible to predict the course of children's future emotional, cognitive, behavioral, and social development (Graziano, Reavis, Keane, & Calkins, 2007; Measelle, John, Ablow, Cowan, & Cowan, 2005; Miller et al., 2006; Shaw, Lacourse, & Nagin, 2005; Smith, Calkins, Keane, Anastopoulos, & Shelton, 2004). Other studies have shown that early educational intervention based on evaluation and proper treatment of preschool children may enhance their future development, and that this progress is associated with better family relations and cultivation of the children's abilities (Beauchaine, Webster-Stratton, & Reid, 2005; Reid, Webster-Stratton, & Hammond, 2003; Shaw, Dishion, Supplee, Gardner, & Arnds, 2006; Webster-Stratton & Reid, 2003, 2006; Webster-Stratton, Reid, & Hammond, 2004). All these studies have posed fundamental questions of concern to the education system, which its representatives presented to the Investigating Panel. They include the following: What benefit can be derived from early evaluation? What difficulties does it entail and what harm can it cause? How can such young children be evaluated properly? Should the evaluation be based on countrywide or universal norms, or should we make do with the impressions of educators, which—though subjective—are based on extensive, in-depth acquaintance with each child over time in various respects? Is early evaluation liable to cause early (and perhaps erroneous) labeling of children?

2. Goals and Procedure

The Investigating Panel was asked to do the following:

1. Survey the needs of the education system in terms of screening, diagnosis, assessment, and evaluation of typical children in preschool through third grade (ages 3–8)
2. Survey the normative evaluation tools used today in the education system, the purpose of each tool, the information it provides (screening, diagnosis, evaluation of functioning), the age range for which it is intended, how it is used, who uses it, what training is required to use it, and to whom the information gathered with it is given
3. Identify the components of the disparity between real and ideal in the education system, i.e., needs that are being met and needs that are not being met
4. Recommend ways of broadening knowledge, developing appropriate tools, and improving teaching, diagnosis, assessment, and evaluation

The Investigating Panel added another goal: to examine diagnosis, assessment, and evaluation in early childhood in Israel from the theoretical and practical standpoints, focusing on three main aspects of development: (a) socio-emotional adjustment; (b) language and linguistic literacy skills; (c) mathematical skills. This goal was added because without it the panel's work lacked a framework.

These objectives have several implications. Diagnosis, assessment, and evaluation are processes that sometimes differ from one another in their essence, their goals, their inputs, and their outcomes. The Investigating Panel was asked to discuss as many of the processes as possible, with emphasis on their common aspects and attention to the uniqueness of each of them in order to cover all the needs of the education system and to provide as broad and in-depth a picture as possible.

The target population for diagnosis, assessment, and evaluation was defined as children in preschool through second grade. The idea was to follow evaluation of the children over the developmental spectrum—from the beginning of institutionalized formal education in preschool to the end of the acquisition of basic skills in second grade, when Israeli pupils take the GEMS (Growth and Effectiveness Measures for Schools) tests. The Investigating Panel was also asked to look at third-graders in order to see the ramifications of the GEMS tests and follow up on the lessons learned from their results; however, because the switch from learning vocalized Hebrew in grades 1 and 2 to learning unvocalized Hebrew in grade 3 requires mastery of advanced language and linguistic literacy skills that have not yet been acquired, the panel decided to study only children aged 3–8 in preschool through second grade. This is a heterogeneous age group in terms of the children's level of development, abilities, and mastery of various skills. Therefore the decision was made to divide it into four subgroups: pre-nursery (ages 3–4), nursery school (ages 4–5), compulsory preschool (ages 5–6), and grades 1–2 (ages 6–8).

The Investigating Panel recommends that diagnosis, assessment, and evaluation in each age group and subject be development-appropriate. The development of appropriate tools for ages 3–5 is especially important because for budgetary reasons there is a tendency to focus only on children in compulsory preschool and older. Regarding the Arab, Druze, and Bedouin sectors, there is insufficient research-based knowledge to determine the desired age range. At present, Arabic-speaking children in those sectors are tested in second grade, like the Hebrew-speakers.

Another important implication of the objectives set for the investigation is the need for communication with leading officials in the Ministry of Education. In order to describe the present situation and identify the diagnostic, assessment, and evaluation needs of the education system, the Investigating Panel remained in close contact with a wide range of Ministry officials, gathering data and obtaining information. For instance, a meeting was held with a wide range of representatives of relevant units in the Ministry, who described the diagnostic, assessment, and evaluation needs of their units, as well as the processes and tools that they currently use to meet those needs. The panel members also consulted with senior personnel in the Ministry's Psychological Counseling Service and Department for Pre-School Education on the socio-emotional development of children; with diagnosticians, inspectors, instructors, and representatives of the Department for Pre-School Education on mathematical thinking; and with representatives of the Department for Elementary School Education on language acquisition and linguistic literacy skills. The information obtained from these sources included study materials and relevant tools for diagnosis, assessment, and evaluation. This collaboration was important and productive.

The work of the Investigating Panel differed from that of other committees in the education system in general and those formed by the Initiative for Applied Education Research in particular:

- The panel was asked to answer, professionally and quickly, all the questions stemming from the delineation of its tasks: it was supposed to survey the present situation, describe the disparity between real and ideal, and formulate recommendations for ways of improving diagnosis, assessment, and evaluation of children aged 3–8 in the education system.
- The Investigating Panel functioned as a “pre-committee.” It was stated that the panel's recommendations might refer, *inter alia*, to the introduction of various learning procedures such as initiating and organizing professional workshops, producing peer-reviewed summaries of relevant knowledge, conducting programs to train professionals and hone their skills, and convening expert committees.

The chair and nine members of the panel are experts in content and/or the main aspects of development discussed by the Klein Committee to Review Modes of Education in Early Childhood

(language and linguistic literacy, mathematics, and socio-emotional adjustment) or are experts in methodology and psychometrics. This multidisciplinary composition made it possible to combine issues related to development and fields of content (e.g., whether the investigation should encompass all fields of knowledge and skills relevant to the age group in question) with methodological and psychometric issues derived from a critical analysis of the quality of the diagnostic, assessment, and evaluation tools used in the education system and their use in practice (e.g., the validity and reliability of the tools, interpretation of the findings, and decisions derived from their application). Future panels that develop indicators should also include experts in other fields related to education and development in early childhood (e.g., the arts, science, and physical education).

Even though the Investigating Panel was authorized to focus on normal children, it also looked at exceptional populations (slow learners and the gifted and talented). This attention to exceptional populations stemmed from the diagnostic, assessment, and evaluation needs of the Israeli education system that emerged in conversations between the panel members and representatives of the education system.

3. Major Milestones in the Work Process

In preparation for the work of the Investigating Panel, the Initiative for Applied Education Research commissioned two professional surveys as background papers. The first survey mapped the tools in Hebrew that are used in Israel for diagnosing, assessing, and evaluating children aged 3–8 in various respects (motor, cognitive, emotional, and social functioning). It also specified who is in charge of developing such tools, who is to use them, and in what contexts (see Appendix C). The second survey dealt with policy regarding diagnosis, assessment, and evaluation of children aged 3–9 and testing procedures in five comparison countries (the UK, the United States, the Netherlands, Finland, and France). The two surveys provided the panel members with information about diagnostic, assessment, and evaluation practices in countries with different worldviews; the essence of these procedures; their purpose; suitable tests; and ways of using the tests.

After studying the background papers, the panel met with senior representatives of Education Ministry departments and units, the Pedagogical Administration, and the Pedagogical Secretariat (the departments and units are listed in chapter 1) to survey the needs of the Ministry with respect to diagnosis, assessment, and evaluation. The information obtained at this meeting from Ministry personnel in charge of implementing Ministry policy and making decisions in their areas of responsibility served as the primary frame of reference for the work of the Investigating Panel. The meeting, presided over by the director general of RAMA, enabled the panel members to become

acquainted with the officials in charge of diagnosis, assessment, and evaluation in the education system, to gain a preliminary impression of the tools that they currently use, and to understand what tools are lacking.

In the course of its deliberations, the Investigating Panel split up into four sub-teams by area of expertise: a sub-team on language and linguistic literacy, a sub-team on socio-emotional development, a sub-team on mathematics, and a sub-team on methodology. The discussions of the Investigating Panel were held in meetings attended by all the members and in meetings of the sub-teams, which sometimes brought in outside consultants, especially officials in the education system. The meetings of the entire panel were devoted to discussing the intersection of the sub-teams' conclusions and decisions and making generalizations based on them, especially with respect to issues common to all the areas examined.

The Investigating Panel conducted a comprehensive, detailed survey of needs in regard to mathematics. This unique survey was required because it turned out that the goals set for evaluating mathematical ability in early childhood were stated too vaguely, and that the available research-based knowledge is extremely limited. The math sub-team held numerous meetings and consultations with representatives of math education in preschool and the lower grades of elementary school and with diagnosticians and researchers who work in this field.

The meetings of the entire Investigating Panel and the sub-teams yielded the recommendations described below.

4. Frame of Reference and Professional Considerations

The Investigating Panel's frame of reference included three different categories with respect to diagnosis, assessment, and evaluation of children aged 3–8 in a multicultural society: (1) the individual; (2) the system; (3) evaluation of special populations and evaluation in a multicultural society.

4.1 Identification and Diagnosis of Exceptional Individuals

This document focuses on identifying and diagnosing exceptional individuals. The diagnosis includes an evaluation of the individual in order to identify, as early as possible, outstanding talents, difficulties, obstacles, deficits, and exceptional functioning of children with respect to aspects of learning and behavior. The evaluation of the individual addresses knowledge, skills, competencies, perceptions, and attitudes of various sorts (motor, cognitive and meta-cognitive, emotional, behavioral, and social). This evaluation is used for diagnosis, assistance, treatment, placement (adaptation of the learning environment), and follow-up. An evaluation of this sort is

generally conducted in two stages: preliminary screening to identify exceptional pupils and thorough diagnosis of them (both issues are discussed below).

4.2 Evaluation of the Progress of the Education System over Time from a Developmental Perspective

Change is frequent in the education system, and the changes and their ramifications over time have to be evaluated. A systemic evaluation generally looks at various dimensions and aspects of the education system and its components (preschool/class, educational institution, district, local authority, type of school [State, State-Religious, etc.], regular/special education, sector); is used for monitoring, control, and policymaking; and serves as a fundamental component of decision-making processes. The purpose of a systemic evaluation is to provide policymakers and decision-makers in the Ministry of Education with a status report on the system in general and to facilitate the monitoring of trends and changes over time in order to identify aspects that require intervention and improvement. Such an evaluation also provides information about processes and outcomes and about the various factors that affect them. The sources of information for a systemic evaluation are diagnoses, assessments, and evaluations of individuals and special activities to gather specific information about the system.

4.3 Evaluation of Special Populations and Evaluation in a Multicultural Society

This area includes evaluation of special-needs pupils (e.g., the learning-disabled and the gifted and talented) and of different cultural groups (e.g., Arabic-speakers, the ultra-Orthodox, and immigrants). The diversity within and between these groups of children is large and requires special, equal, and fair attention irrespective of race, ethnicity, culture, language, gender, religion, and physical or mental ability. It turns out that in this area, which is based on the right of all pupils to equal opportunities to develop and learn, the shortage of standardized tools is especially severe.

5. Survey of the Needs of the Ministry of Education

The Investigating Panel thoroughly reviewed the tools and processes currently available to the education system for diagnosing, assessing, and evaluating children in preschool through second grade. This review included an examination of their psychometric validity and reliability, fairness, and suitability for the Israeli education system and considered their potential for diagnosing, assessing, and evaluating typical and exceptional individuals, as well as programs and tracks in the education system as a whole.

The in-depth examination showed that there are very few tools for diagnosis, assessment, and evaluation in Israel that have valid, reliable countrywide norms and very few tools designed for

testing individuals with special needs in a multicultural society (see Appendix B). Consequently, the panel decided to focus first and foremost on the Mabatim tool developed in recent years in Hebrew and Arabic by the Department for Pre-School Education to test the functioning of preschool-age children. The panel sought to determine how suitable this tool is for diagnosis, assessment, and evaluation; whether it provides valid, reliable information about children's development and achievements; and how it can be improved.

The methodological principles used to examine the few valid, reliable tools available in Israel for diagnosis, assessment, and evaluation in early childhood are described extensively in the recommendations for the development and use of evaluation tools (chapter 3 below).

Chapter 1

Survey of Needs of the Ministry of Education with Respect to Diagnosis, Assessment, and Evaluation in Preschool through Second Grade

This chapter sums up the discussion that took place at the Investigating Panel's meeting with representatives of the Ministry of Education and the National Authority for Measurement and Evaluation in Education (RAMA). The goals of the meeting were to study the needs and become familiar with existing tools for diagnosis, assessment, and evaluation of children aged 3–8 in Israel, to map them, and to identify disparities between real and ideal in these realms. The Education Ministry representatives included officials from the Pedagogical Secretariat, the Department for Pre-School Education, the Department for Elementary School Education, the Department for Curricula, the Psychological Counseling Service, the Department of Education and Welfare Services, the Department for Talented and Gifted Students, the Department for Special Education, and the Immigrant Absorption Unit. These representatives were asked to describe their units, their areas of responsibility, the needs with respect to diagnosis, assessment, and evaluation of children aged 3–8, the tools and processes available to them for meeting these needs, and the needs that are not being met. The participants in the meeting handed out background material for perusal and clarification before and during the meeting.

The information provided by the Education Ministry representatives is presented below.

1. Needs of the Psychological Counseling Service

Tools used by psychologists: In their work for the Psychological Counseling Service, the psychologists use standardized psychological tools—intelligence tests in Hebrew that have been standardized for the Israeli population (WISC-R-95, K-ABC, WPPSI) and the Elul test kit in Hebrew and Arabic to evaluate cognitive functioning, as well as psychological tools without Israeli norms that have been translated into Hebrew but are based on standardization for a Western population. The latter include cognitive tests to identify difficulties and examine school readiness (Shatil, Maslag, and Ituran), socio-emotional and behavioral tests (Connors and Achenbach questionnaires, CAT Roberts), and perceptual-motor tests (Bender-Gestalt, Beery, and MAP). A psychological evaluation includes observations, an interview with the teacher, and screening tools

used by the teacher (Mabatim, conversation with parents, conversation with and interview of the child, RTI [Response to Treatment Intervention]).

Tools used by the Department for Pre-School Education: The Department for Pre-School Education has created an observation-based tool known as Mabatim for use by preschool teachers. Mabatim, developed and applied in Hebrew and Arabic, enables teachers to work dynamically with all the children in the preschool and to monitor aspects of normal development and functioning. It can identify temporary or lasting difficulties and makes it possible for teachers to tailor interventions addressing linguistic, motor, cognitive, and behavioral problems. Preschool teachers are trained in the use of Mabatim by panels from the Psychological Counseling Service, child psychologists, child counselors, and instructors from the Department for Pre-School Education.

Tools used by teachers under the supervision of the Learning Disabilities Unit: A team from the University of Haifa recently developed a tool to identify learning disabilities and learning difficulties as part of the Elul project (a test to diagnose learning disabilities). The Elul tests have countrywide Israeli norms and have been validated separately for Hebrew-speaking and Arabic-speaking populations. The Hebrew version was provided to the Learning Disabilities Unit of the Ministry of Education in August 2007. The kit contains a user's manual, which gives detailed descriptions of a series of tests, including tasks, scoring methods, and ways of interpreting the scores. The manual also summarizes case studies and presents psychometric properties of the tests (indices of validity, reliability, etc.). The tests are designed to be administered to entire classes in grades 1–9. The series of tests includes:

- Native language test: several subtests of listening comprehension, reading comprehension, reading of real words and nonsense words, spelling and graphomotor skills, and exposure to the written word
- Foreign language test: an English test for the Jewish and Arab sectors, including subtests on reading of real words and nonsense words, vocabulary, reading comprehension, listening comprehension, and spelling
- Math test: a test of early numeracy, sequential verbal processing of numbers, comparison of sizes using numbers, visual comparison of quantities, visual (not numerical) processing, motor speed, and mathematical discourse
- A test of thinking/intelligence: a test that examines verbal as opposed to non-verbal processing, vocabulary, logic, and various cognitive abilities (with the help of mazes, cubes, etc.)
- An attention and concentration questionnaire: a questionnaire for homeroom and subject teachers that examines organization abilities, attention, and activity (e.g., hyperactivity)

Unmet Needs

Psychological Counseling Service

1. There is an urgent need for standardization of a new IQ test for Israeli children. The test norms standardized in the past are outdated, and there is no standardized test for the Arab or ultra-Orthodox sector.
2. A standardized diagnosis of developmental levels is needed in various realms: intellectual, cognitive, emotional, social, and behavioral.

Recently the Psychological Counseling Service formed an “off-the-shelf test” team to meet these needs and review the tests, especially the IQ tests.

Learning Disabilities Unit

1. A tool is needed to test aspects of functioning for early identification of three-year-olds at educational risk.
2. A tool is needed to evaluate the development of spoken language skills as an indicator of their quality and a predictor of the acquisition of reading ability (as opposed to tests of reading skills for grades 1 and 2, which already exist).
3. A behavioral screening tool is needed for use by various people in the system. This tool should focus on attention and concentration problems, behavioral problems, social withdrawal, and emotion regulation and effortful control.

2. Needs of the Department of Education and Welfare Services

The needs are great and there are few means available for meeting those needs. Specifically, what is required is as follows:

1. Criteria for selecting programs to implement
2. Appropriate tools for monitoring intervention programs in use

3. Needs of the Department for Pre-School Education

Use and evaluation of the preschool curriculum are intertwined. Evaluation is an ongoing process that tracks progress in achieving short-term and long-term goals and is carried out by the educational staff.

Every evaluation in early childhood must take into account the following aspects:

- The wide range of developmental norms
- The critical impact of the home environment and family culture

- The impact of the presence of observers on the observed behavior (as observed behavior changes frequently due to internal and external factors)
- Potential ethical problems, such as early labeling and the effect on the family, that may result from the evaluation process

Documentation and evaluation are currently carried out by the preschool teachers. The teacher's evaluation is based on information gathered in various situations with the help of tools such as observations, analysis of products produced by the children, documentation of conversations with and among the children, documentation of free play, documentation of conversations with parents, and analysis of photographs and video recordings of preschool activities. The main diagnostic tool used to identify problems and difficulties is Mabatim, which looks at all aspects of development based on systematic observation of children in the natural preschool environment. The goals of the evaluation are to monitor the children's development and learning in order to plan methods of teaching/learning and to design an educational environment; to identify problems and difficulties that require specific intervention; to involve the children's parents and report to professional agencies; and to identify children in need of professional diagnosis and intervention (by the Psychological Counseling Service, medical professionals, and social services). A screening tool is currently needed for identifying children having special difficulties; Mabatim was not designed for this purpose. There is some ambivalence about designing such a screening tool because the department, at present, is poorly equipped to use the information to help the children and their parents. In this situation, the feeling in the department is that the screening would label children without helping them in any way.

4. Needs of the Department for Elementary School Education

Diagnosis, assessment, and evaluation in elementary school are part of the teaching/learning culture and are based on the routine gathering of data on schools, classes, and individuals. These processes focus on evaluating the degree to which curricular standards for language education are being met (Ministry of Education, Culture, and Sport 2003). Their purpose is to identify difficulties as early as possible, without excessive labeling and classification of pupils. The assessments and evaluation activities are carried out by the teachers, who teach and evaluate the children simultaneously. The homeroom teacher, and not other teachers, works with slow learners. The children and their parents are perceived as partners in the diagnostic process, and the teachers are committed to documentation, follow-up, and evaluation. The Department of Education and Welfare Services also helps out with the diagnostic, assessment, and evaluation activities.

The stages of diagnosis, assessment, and evaluation are screening, individual diagnosis, creation of a reader profile in the wake of the diagnosis, design of individual and group work plans, use of the plans by educators in the regular learning environment, documentation, control, follow-up, and evaluation. The teachers use diverse tools, most of which are based on their discernment and judgment, although some are based on countrywide norms. The tools, which may or may not come with scoring rubrics, enable the teachers to carry out individual, group, or class evaluations of children of various ages.

The following tools are available to teachers: class survey; first-grade reading and writing test; conceptualization of the text of a book; dynamic diagnosis with the help of a narrative text (individual); dynamic diagnosis with the help of a scientific text (individual); tests to evaluate various genres of writing (individual and group); and various criterion-referenced or norm-referenced tools—From Aleph to Tav (a test to diagnose reading difficulties); the Ma’akav kit (an up-to-date method for surveying reading and writing abilities); a portfolio of reading and writing assignments; the Elul kit for each age group; and GEMS tests.

Unmet Needs

1. Continued cooperation between academics and practitioners with respect to theory and practice
2. Uniform criteria for the evaluation tools used by practitioners and external evaluators
3. Training to take an intelligent perspective on data-gathering
4. Normative tests for the Jewish sector in unaddressed fields of content
5. Norm-referenced tools for the Arab sector
6. Alternative evaluation methods and evaluation tools (“off-the-shelf tests”) that can be selected: norm-referenced tests for monitoring the level of spoken conversation; tools to examine the extraction of meaning from nonfiction and narrative texts; tools for examining writing skills for communication purposes; tools for identifying reading and writing skills among children entering first grade
7. Professional knowledge of evaluation based on the various fields of knowledge, such as tests of comprehension and vocabulary
8. Avoidance of inundation with unnecessary tests and diagnoses

5. Needs of the Department for Arab Education

The tools available in the Arab sector for diagnosis, assessment, and evaluation in preschool through second grade are the GEMS test in Arabic for second-graders, which tests reading, writing of words and sentences, reading comprehension, and written expression; the Elul test kit (Arabic-

language version) for first- and second-graders; and the Mabatim program for identifying learning-disabled children from age 4 for the purpose of appropriate treatment. The department also has irrelevant tools that do not meet its needs. Most of the tools have been translated from Hebrew into Arabic without being adapted for Arab culture. Unique tools are needed that are tailored to the Arab population and Arab culture. It is necessary to examine whether the age range included in the definition of early childhood (3–8) is appropriate for the Arab, Druze, and Bedouin sectors.

Unmet Needs

1. A comprehensive test of language-related achievements for various ages: listening, speaking, reading, writing, and linguistic knowledge
2. A test to monitor the acquisition of reading and writing skills in first grade, tailored to the Arab sector (one already exists in the Jewish sector)
3. Solutions to problems other than those pertaining to diagnosis, assessment, and evaluation:
 - (a) Curriculum units must be adapted to the Arab sector. The study materials currently used in elementary schools are written in standard Arabic, whereas classes (including math classes) are taught in the spoken language. Because there are major differences between spoken and standard Arabic, the language of instruction in all subjects other than Hebrew and English should be standard Arabic.
 - (b) The elementary-school curriculum is very outdated. The Department for Curricula reports that when a new curriculum is being prepared, it is difficult to find researchers who can help write curriculum units in Arabic when these units are originally written in Arabic (and not translated from units developed in Hebrew)
 - (c) Until now curriculum units have been translated from Hebrew and adapted for the Arab population. Sometimes these units are outdated. The result is a disparity between real and ideal. Curriculum units should be written for the Arab population in Arabic from the start.

6. Needs of the Department for Talented and Gifted Students

Talented and gifted students are identified by means of group psychometric tests that examine general academic ability only. These tests are used as a standardized tool for initial screening. The tests are given simultaneously in Hebrew and Arabic (the tests for the Arab sector were composed specially in Arabic). Success on these tests is a criterion for acceptance into special programs for this population. The availability of valid, reliable screening tools is not merely a technicality but a fundamental ideological issue because they are used to identify students who are eligible for additional enrichment. The extra investment also creates especially high expectations.

Unmet Needs

1. A research study is needed that can come up with valid, reliable criteria for identifying gifted students and determining their unique individual traits with respect to various fields of knowledge (language, science and technology, art, etc.). Such criteria should make it possible to refer potentially outstanding students to intervention programs.
2. The tools for identifying gifted students should be objective and universal. At present, schools decide for themselves (usually based on teachers' impressions) which students to refer to special programs for the gifted. This screening process is too subjective and should be replaced with an objective, universal process for selecting gifted students from among all the students in the education system.

7. Needs of the Department for Special Education

Diagnosis, assessment, and evaluation for the purpose of referring children with special needs to special education are carried out by teachers, therapists, guidance counselors, psychologists, and social workers. The evaluation is holistic and addresses all aspects of functioning: motor, cognitive, behavioral, linguistic, organizational, and academic. Each diagnostician uses his/her own tools and presents his/her findings regarding each child in a staff meeting at which it is decided, based on all these findings, whether special treatment is needed. If it is, objectives for treatment are set, and an individualized plan of action is devised.

The tools currently used for evaluating special-needs children (some standardized and some not) are based on the stages of normal development. These tests (Hatzav, the partially distributed Shvilim kit), which were designed by local teams, are specially tailored to these students and they examine several aspects for various purposes:

1. An evaluation of overall functioning for the sake of an individualized treatment and support plan. Only some of the tests of aspects of functioning are standardized. Often non-standardized tools are used that were devised by the therapeutic agency on its own.
2. Diagnoses for the purpose of referral to committees as required by law. These are based on an evaluation form filled out by the educational staff, along with documents from professionals (didactic, psychological, audiological, and other diagnoses) that describe the disability and level of functioning.
3. Diagnosis of the level of functioning in order to determine the kind of support needed
4. Evaluation of language-related achievements, e.g., level of communication, use of language and speech
5. Evaluation of achievements in mathematics

6. Evaluation by an occupational therapist
7. Evaluation by a physical therapist
8. Didactic diagnosis
9. Diagnosis of children with sensory disabilities

Unmet Needs

1. Diagnostic tools and tests that have been standardized on the countrywide level and tailored to various populations in special education
2. Criteria and standards for diagnosing the spectrum of disabilities for the purpose of appropriate screening of special-needs children. The main difficulties with diagnosis today are among young children, children with moderate and profound retardation, and children with complex disabilities.
3. Standardized, valid, and reliable diagnostic tools for comprehensive testing of aspects of ongoing development in the Arab sector. Such tools must be designed in Arabic (and not translated from Hebrew).
4. Greater use of existing tools for occupational and physical therapy
5. A reliable, standardized tool for diagnosing language-related learning disabilities. The tools available today are not adequate. In order to develop such tools, the panel recommends liaising with academics and relying on an extensive database of linguistic components in Hebrew.
6. Further development of diagnostic tools for determining the type and level of support needed, based on the level of functioning
7. Policy on the following issues: (a) When should the present diagnostic, assessment, and evaluation tools (adapted to the target populations) be used and when should special tools be developed? (b) How should tools be adapted without changing their essence? (c) How should norms be set for groups of students on a similar level of functioning or for age groups? (d) How should evaluations be expressed on the report cards of mainstreamed children and children in special education?

8. Needs of the Immigrant Absorption Unit

The Immigrant Absorption Unit implements several educational intervention programs for integrating immigrant children aged 3–8 in Israeli society. All the programs are evaluated by RAMA:

- Guidance programs for parents and their children: HATAF (Home Activities for Toddlers And their Families) for ages 1–3, and HA'ETGAR (a home instruction program for parents of preschoolers) for ages 3–6
- Programs for social integration and development of Jewish and Zionist identity in order to overcome a sense of alienation among immigrant children
- In-service programs and facilitation for teachers of immigrant children
- The PACT (Parents and Children Together) program for greater parental involvement in the educational process in elementary schools. The program is evaluated by the Brookdale Institution.
- Outside examiners from RAMA administer a test at the end of the stage of intensive Hebrew classes, before the children join their regular classes, to evaluate the quality of absorption and readiness to enter regular classes.

Unmet Needs

1. Cooperation should be established among everyone involved in the unit's activities and agreement should be reached as to who performs the diagnoses, assessments, and evaluations.
2. A reliable questionnaire is needed to obtain information about immigrant children's academic knowledge and proficiency in Hebrew when they arrive in Israel, before they start school in their new country.
3. A unique tool is needed for diagnosing the language skills of immigrant children before they enter first grade. Immigrant children who do not speak Hebrew at home enter first grade with substantial deficits and disparities. Such a tool must address the linguistic and cultural diversity among immigrants from different countries (such differences even exist among immigrants from different countries in the former Soviet Union).
4. A tool is needed to evaluate the work of teachers of immigrant students.
5. A tool is needed to diagnose learning disabilities among immigrant children. Special care should be taken to ensure that the diagnosis is not influenced by the child's difficulties with the Hebrew language.
6. A system is needed to evaluate educational materials for teaching reading and the Hebrew language to immigrants, based on the children's country of origin and age.
7. Israeli academia should have a chair in research on teaching Hebrew as a second language. The theoretical and practical results of systematic research in this field can improve the absorption of immigrant children in Israeli schools.

8. In addition, there are substantial lacunae with respect to immigrant absorption that have nothing to do with diagnosis, assessment, and evaluation.

9. Conclusion

The reports of Education Ministry departments indicate that there is high, across-the-board awareness of the immediate need for development of normative kits for diagnosis, assessment, and evaluation of children aged 3–8 in all subjects. There are substantial lacunae with respect to testing typical students, gifted and talented students, and special education students on the individual and systemic levels, and a suitable solution is needed with respect to evaluation in a multicultural society. Furthermore, the IQ tests currently used in the education system, which have valid countrywide norms, are outdated and have been standardized only in Hebrew.

Chapter 2

Real versus Ideal: Implications of Evaluation Methods and Specific Recommendations

This chapter will present the Investigating Panel's recommendations regarding socio-emotional development, language and linguistic literacy, and mathematics. We will explain how the panel assessed the present situation (following conversations and meetings with various people in the system) and will describe the implications of this situation and the recommendations for further research and development in each specific area.

1. Socio-Emotional Development

Healthy socio-emotional development in children contributes to cognitive development and efficient learning and is conducive to adjustment to school. The socio-emotional development of young children entering the education system varies widely. Unless the system makes a special investment in the progress of children with socio-emotional difficulties, these difficulties are liable to intensify as the children get older, especially if they are growing up in disadvantaged conditions and an unsupportive home and community environment.

Healthy socio-emotional development is not only a condition that enables young children to learn and develop in preschool and school, but also an objective that the preschool or school is supposed to achieve. This perspective on socio-emotional development has implications for diagnosis, assessment, and evaluation of children and of educational settings, including the socio-emotional climate that prevails in them. Healthy socio-emotional development enables young children to cope with separation from parents, to trust their teachers, to internalize rules and obey instructions, to communicate with other children their age, to form friendships and partnerships for play and learning, to regulate emotions in various situations, to focus their attention, to persist in tasks, to diversify their activity and show flexibility, to cooperate with other children, to empathize with others, to develop a positive self-image, to be proud of their personal achievements, and more. Emotional and social difficulties are manifested in a wide range of behaviors, some of them externalized (e.g., aggression) and some internalized (e.g., fears and anxiety). The course of children's socio-emotional development is individual and depends on age and context. Consequently, diagnosis, assessment, and evaluation of this process are especially complex.

1.1 Screening

1.1.1 The Present Situation

At present, the education system does not have valid, reliable tools that have been standardized in Israel and can be used for diagnosing, assessing, and evaluating the socio-emotional development of young children. Over the years, tools for evaluating children's readiness for first grade have been developed in various places in Israel, and the Psychological Counseling Service currently has a committee that is gathering information on these tools. The Investigating Panel was not given these tools, partly because they were never developed beyond a preliminary level that is far from meeting the present needs. The only tool developed on the countrywide level and currently being disseminated in the education system is Mabatim, which is based on observations and impressions of the preschool teacher and has not yet been validated (see below).

1.1.2 The Ideal Situation

The education system needs valid, reliable evaluation tools that will make it possible to identify children with special socio-emotional needs and to develop intervention programs that meet the needs of these children at as young an age as possible. Evaluations should be based on a developmental survey model (e.g., like the model used in well-baby clinics for diagnosing infants and young children) and should avoid labeling the children by means of conclusive diagnoses. It should evaluate children's performance and their emotional, behavioral, and social functioning along a normative spectrum. A valid, reliable evaluation of these aspects will enable the education system (1) to identify children at risk of developing behavioral problems, attention and concentration problems, and emotional and social distress; (2) to diagnose these children early on when developmental flexibility is still fairly high, before problematic patterns become fixed in place; (3) to implement preventive intervention measures tailored to the needs of individual children and the various groups of children. A tool is also needed for mapping geographic regions with special needs, evaluating the progress of the education system in this regard over time, and dealing with a multicultural society and evaluation of special populations.

Mabatim, the tool commonly used today in preschools to evaluate children's functioning, provides preliminary information about the children's socio-emotional development and their environment and can be incorporated in the screening procedure. The Department for Pre-School Education developed Mabatim and distributed it as an observation-based tool. The use of Mabatim includes training preschool teachers to conduct detailed observations of various kinds of functioning in several areas of activity. In addition to Mabatim, the Ma'agan intervention program helps preschool teachers adapt their work to children with special needs after having identified

them by observation. It seems that this is a positive initiative based on systematic work by the Ministry staff. The developers of Mabatim gathered data and devised a tool that is receiving recognition and being used countrywide. However, they did not consult with researchers and academics, and the tool has not been validated by research. Therefore interpretations of observations based on Mabatim are largely subjective impressions, and as things stand at present, the tool is not valid for screening (for a detailed discussion of the development of Mabatim, see chapter 3).

We recommend that the education system become familiar with the SDQ (Strengths and Difficulties Questionnaire), a screening questionnaire designed to identify children suffering from behavioral and emotional disorders and in need of further treatment. This relatively short (one-page) questionnaire contains questions about the child's positive and negative traits, and no special skill is needed to administer it. The validity and reliability of this questionnaire are supported by numerous studies around the world. It has been translated into more than sixty languages and widely use as a tool for identification, diagnostic assistance, research, and evaluation of the results of treatment. An SDQ-based study conducted in coordination with the Ministry of Health can provide Israeli norms for school-age children, but further research is required in order to create Israeli norms for the use of the SDQ with ages 3–9.

1.2 Diagnostic Evaluation

1.2.1 The Present Situation

The education system currently has two types of diagnostic tools: (1) projective tests with disputed validity that are not appropriate for widespread use as standard diagnostic tools for reasons to be explained below; (2) structured tools that have been validated abroad (with the exception of the Child Behavior Checklist [Achenbach & McConaughy, 2003], which has also been validated in Israel) and are used in research. The Child Behavior Checklist (CBCL) is associated with Achenbach's diagnostic system, and the validated Hebrew version is an old one. The Hebrew version currently in use does not have up-to-date Israeli norms and has not been adapted for the multicultural Israeli society.

Our impression is that the tools currently available in Israel and used by psychologists for diagnosing, assessing, and evaluating socio-emotional difficulties are mainly projective tests such as the Children's Apperception Test (CAT) and House-Tree-Person (HTP), and sometimes Rorschach as well. It seems that the main reason for the widespread use of these tools is the absence of a significant alternative (as stated, there are structured evaluation tools such as CBCL but they lack up-to-date Israeli norms) and their association with the psychoanalytic tradition that to

this day influences the training of clinical psychologists. Even though most research-backed intervention programs for children and teenagers focus on cognitive, behavioral, and family issues and are not associated with the psychoanalytic tradition, educational psychologists are still largely trained to diagnose in accordance with the standard model used by clinical psychologists.

A wide variety of structured tools are available abroad, but the few that have been translated into Hebrew do not meet the needs. For instance, the norms for Achenbach's diagnostic system relate to an old version of the tool; the Connors questionnaire for evaluating attention and concentration problems, which is currently distributed commercially in Israel, has only American norms. Moreover, the cost of using the tools distributed by commercial firms (which have translated them into Hebrew) is high and is a burden on psychological services that wish to use them.

1.2.2 The Ideal Situation

The following factors support the recommendation to consider changing the present tradition in Israel regarding diagnosis, assessment, and evaluation of young children:

1. We cannot ignore the question marks regarding the diagnostic validity of projective tests. To put it mildly, the research findings do not support continued widespread use of projective tests for diagnostic evaluation. There are, however, structured tools with good reliability whose diagnostic validity is supported by research.
2. Diagnosis by means of projective tests is very time-consuming. Given the heavy workload of educational psychologists, which greatly limits their ability to engage in intervention activities, and given that most of the interventions are short-term, it would be best to use diagnostic tools that take less time. The evaluation tools whose validity is supported by research are economical and efficient in terms of personnel. Diagnosticians can use them to gather information relatively quickly from educators, parents, and children and to produce a computerized report. Thus they can save time and obtain valid, reliable results while comparing the data on the individual with norms adjusted to his or her reference group.
3. It should be stressed that we are not rejecting projective diagnostic tests entirely. It is clear to us that projective tests and observations can be used to get to know the child better one-on-one and to gather narrative material alluding to the child's socio-emotional difficulties and the way he/she sees the world in order to help form a relationship and embark on therapeutic intervention. We believe that even when the education system decides to use structured diagnostic tools as its default, there will be situations in which projective tests will be worth using as well.

We recommend developing at least two tools with multicultural Israeli norms that meet criteria for validity and reliability and will help the educational psychology system streamline the diagnostic process: one tool should address behavioral, emotional, and social problems (areas covered by Achenbach's diagnostic system and similar systems); the second tool should be designed specifically for diagnosing ADHD. The two tools should focus on structured, efficient gathering of information from parents and educators.

In addition, we recommend that the "off-the-shelf test" team in the Department of Psychology of the Psychological Counseling Service look into structured tools used abroad, both those that have been translated into Hebrew and those that have not yet been translated. These include diagnostic tools that are used extensively in educational psychology in the United States, such as the Devereux Early Childhood Assessment (DECA) and the Behavior Assessment System for Children (BASC). Like Achenbach's diagnostic system, these are questionnaires, several versions of which have been validated for preschool- and school-age children, and they make it possible to gather information from teachers and parents. We recommend looking into the possibility of obtaining the rights to translate these questionnaires, adapt them for Israel, and develop Israeli norms. The Ministry will have to develop or adopt tools that are not prohibitively expensive to use. This will mean deciding whether to develop new tools suitable for the various segments of the Israeli population or to rely on existing tools from abroad and adapt them to Israeli society.

Tools are also needed for evaluating children's educational environment and its suitability for their emotional-development needs. It may be necessary to devise a questionnaire for preschool teachers, families, and health and social services in order to assess what exists and what is lacking in this regard.

2. Language and Linguistic Literacy

Mastery of the spoken and written language and the ability to read and write are basic conditions for being a part of today's global and technological society, in which human communication, which is based increasingly on the written word, is becoming quantitatively and qualitatively richer. The level of linguistic literacy has far-reaching economic implications because in order for this country to be an integral part of the global market, there must be a high level of literacy among broad sectors of society. Furthermore, the ability to read and write has a clear impact on the level of education of the individual and of society as a whole because the information imparted in school or obtained via media such as the Internet is often presented in written form. Therefore reading difficulties in general and reading disabilities in particular, as well as difficulty in writing texts, are a serious obstacle to acquiring and making use of one's education. Despite the tremendous

investment by education systems in furthering reading instruction in general and resolving reading difficulties in particular, the achievements are often fairly disappointing. All this is true in Israel as well, as shown by comparative international studies such as the Progress in International Reading Literacy Study (PIRLS). In recent years, a concerted effort has been made in various countries, including Israel, to improve this aspect of education. Some examples are the Committee on Reading Instruction, headed by Prof. Rina Shapira; the follow-up committee headed by Prof. Joseph Shimron, which recommended changes in the reading materials used; and the Committee to Foster Literacy in Preschool, headed by Prof. Iris Levin.

The Committee to Review Modes of Education in Early Childhood, headed by Prof. Pnina Klein, focused on education in preschool through second grade (Klein and Yablon, 2007) and recommended starting to work on language and linguistic literacy in early childhood, i.e., at age 3. Because the goals of teaching and learning are closely related to the processes of diagnosis, assessment, and evaluation, implementing this recommendation will make it necessary to evaluate skills and achievements in these subjects from age 3. It should be noted that language and linguistic literacy develop and improve throughout the years of schooling. It is not enough to invest in these subjects in early childhood; nor is it enough to have evaluation tools suitable for these ages only. The effort should continue throughout the years of schooling.

2.1 The Present Situation At present we have few valid, reliable tests with countrywide norms for diagnosing, assessing, and evaluating language and linguistic literacy skills. The ones currently available are as follows:

1. Tests for grades 1–2: These tests are designed to examine reading skills in a class setting and individually and to be administered by the class teacher or a learning evaluator. The class tests include an achievement test for acquiring reading in the first grade (A. Ben-Simon, Shani and Cohen, 2005). This test is given to all first graders by the homeroom teacher and is used to find children who do not have reading skills at the end of first grade; Ma’akav (administered by the teacher), for identifying difficulties in reading Hebrew (Shani, Lachman, Shalem, Bahat and Zeiger, 2006), and Elul (native-language subtest), for identifying and diagnosing difficulties and disabilities in reading and writing Hebrew and Arabic. From Aleph to Tav, a test to diagnose reading disabilities in Hebrew, is administered to individuals by a learning evaluator. These four tests focus on identifying and diagnosing language disabilities among first- and second-graders. The Achievement Test for Acquiring Reading in First Grades was used in the education system in 2006; Ma’akav and From Aleph to Tav were introduced into the system a year or two ago, and Elul was presented to the Learning Disabilities Unit in the Ministry of Education in August 2007 for preliminary review. It is still too early to evaluate the

effectiveness of these three tests. From a theoretical standpoint, they address three reading skills: decoding written language, reading comprehension, and reading fluency. They also examine basic reading skills (phonology, orthography, working memory, and naming). We recommend that the quality of these tests and their ability to meet the needs of the Israeli education system be evaluated on the basis of research. In addition, RAMA has the Achievement Test for Acquiring Reading and Writing in Second Grades (Ben-Simon, Shani and Cohen, 2005). This test is meant to be taken in a group, and can sometimes be used for individual evaluation.

2. Indicators for preschoolers: Mabatim is a multidisciplinary evaluation program developed by the Department for Pre-School Education in the Ministry of Education and used extensively today in preschools. An observation-based tool for preschool teachers, it is used to identify the abilities of preschool children. The rationale for Mabatim is based on research and is consistent with the conclusions of several reports on early childhood. However, this program is not an adequate means of evaluating important aspects of the development of language and linguistic literacy that every preschool teacher should address when evaluating the children's development. The main deficiencies concern description of the children's achievements in early writing, letter recognition, phonology, syntax, grammar, and understanding of higher registers of the language. In order to be able to evaluate these abilities properly using Mabatim, the Investigating Panel recommends relying on the detailed descriptions of the achievements expected of children at different ages in the two curriculum units for preschool and grades 1 and 2: "Foundations of Reading and Writing" (Ministry of Education, Culture and Sport, 2006a) for preschools; and "Language Education: Hebrew—Language, Literature, and Culture" (Ministry of Education, Culture and Sport, 2003) for elementary schools. Different examples should be used for preschools for Hebrew-speakers and preschools for Arabic-speakers. In addition to Mabatim, there are tests of various aspects of language and linguistic literacy. Because these tests lack representative Israeli norms, the Investigating Panel decided not to discuss them.

2.2 *The Ideal Situation*

1. Because each language has its own unique properties, promoting the teaching of a language and instituting an evaluation system require information about the normal development of language and linguistic literacy skills among children in the society that speaks that language. This information must include facts about pupils from different cultures and sectors within the

society. Numerous studies have been conducted in Israel on language and linguistic literacy, especially with respect to Hebrew as a first language; a few studies have also been done on Arabic and/or Hebrew as a second language. It is important to bring together the findings of the various studies in order to develop normative tools for diagnosis, assessment, and evaluation with respect to language and linguistic literacy in Hebrew and Arabic among typical and exceptional populations. The focus should be on diagnosis, assessment, and evaluation of the development of spoken and written language skills, reading, writing, and text composition among children aged 3–8, on levels adjusted for the children’s age, in conformity with the curriculum and taking into account the achievements required in each age group.

2. We recommend that tests of preschoolers be administered to individuals or small groups. The teacher should administer a screening test to children who are having difficulty, and experts should diagnose the type/profile of the difficulty. Preliminary diagnoses of first-graders should be based on the teacher’s impressions or on class evaluation tests such as Ma’akav and/or Elul, which should be given at the end of first grade. Children diagnosed as having difficulty should undergo a more in-depth individual diagnosis by an expert.
3. Diagnosis of children in preschool and the beginning of first grade should focus on subject matter included in the curriculum units for early childhood (“Foundations of Reading and Writing,” Ministry of Education, Culture and Sport, 2006):

Language: We recommend that the teacher carry out a brief, preliminary diagnosis involving vocabulary, syntax, and morphology, and their interrelations.

Linguistic literacy: The diagnosis should be made by the teacher and should address letter knowledge, phonological awareness, early word recognition, and early writing. The Investigating Panel did not reach a unanimous opinion as to diagnosis of concepts about print (e.g., regarding use of the Hebrew version of Clay’s test).

4. Diagnosis of first- and second-graders should focus on the following material (after the children have learned to read):

Diagnoses related to linguistic literacy should focus on skills based on knowledge of the writing system: mastery of the alphabetic code, word-reading level, word-reading fluency, and reading comprehension.

Language-related diagnoses (based on the curriculum unit “Language Education: Hebrew—Language, Literature, and Culture,” Ministry of Education, Culture and Sport, 2003) should focus on listening, speaking, reading and writing texts of various sorts, producing linguistically correct written texts, learning from various types of texts, reading and understanding works of various sorts, and contending with age-appropriate classical

Jewish texts. In this context, it is important to note that children with poor linguistic knowledge, poor literacy skills, and/or limited knowledge of the world need special help with the spoken language in addition to the written language.

Writing-related diagnoses should focus on forming the letters in print and cursive, understanding letter-sound relationships, initial knowledge about representing vowel sounds in particular places in a word, knowing the word-final letter forms and the boundaries of a word, distinguishing between regular and word-final letters and writing them in the appropriate places, leaving spaces between words, and attaching prefixes and suffixes to written words (the letter *heh* as the definite article, *vav* as a conjunction, etc.).

Text-writing diagnoses should focus on the ability to compose a syntactically and grammatically correct sentence and to independently write a simple composition on a selected topic, as well as fluency of writing.

3. Mathematics

Mathematics has traditionally been regarded as a logical field of knowledge that is part of human civilization and as a tool for describing reality in quantitative terms so as to help solve problems in everyday life and in the sciences. In today's scientific and technological society, mathematical applications are needed in many areas of life, and therefore mathematical knowledge is considered an important item in an educated individual's toolbox.

Israel and many other countries ascribe great importance to the development of mathematical thinking at a young age. Various scholars have reported detecting initial indications of mathematical thinking among very young children (just a few weeks old) and claim that children develop a wide range of mathematical abilities (e.g., comparing quantities, counting, dividing into equal parts, distinguishing between numerical and other symbols, and recognizing basic geometrical shapes) at the age of three and even earlier (e.g., Butterworth, 1999; Deheane, 1997; Satlow & Newcom, 1998). Curricula in Israel and other countries therefore recommend developing numeracy in preschool and even earlier. This is also one of the main recommendations of the Committee to Review Modes of Education in Early Childhood, headed by Prof. Pnina Klein ("The development of numeracy should begin in preschool and perhaps even earlier" [Klein and Yablou, 2007]). Great importance is ascribed to mathematical achievement on all grade levels, and knowledge of mathematics is an entrance requirement for institutions of higher education.

The goals of teaching and the mathematical subject matter studied in Israel in each grade are determined by a "subject committee," and diagnosis, assessment, and evaluation on the individual and systemic levels must be based on its decisions. In this context, research should be undertaken to

determine how appropriate the scope and depth of attention to the main subject matter selected by the subject committee are for the various ages from preschool to second grade. International comparisons of curricula indicate substantial differences in goals for each grade. For instance, the Israeli math curriculum states that by the end of first grade pupils should be able to count to 100 and count objects up to 30. In contrast, a recent document by the National Council of Teachers of Mathematics (2006) in the United States recommends that first-graders know how to count and count objects up to 100, whereas a British document written for the National Numeracy Strategy expects children to be able to count up to 10 objects in their first year of school (Department for Children, Schools and Families, 2006). Consequently, in our opinion it is necessary to conduct a study that will provide information about normative mathematical functioning of Israeli children in preschool through second grade in the two main fields of content learned at this age: numbers and operations; and geometry and measurements. The absence of such information calls into question our present ability to diagnose, assess, and evaluate mathematical performance on the individual and systemic levels.

3.1 Needs Related to Evaluation of the Learning of Mathematics in Preschool Through Second Grade: Opinions of Education Ministry Representatives

The math sub-team met with experts in mathematics education, didactic diagnosticians, and representatives of the Ministry of Education to obtain an overview of needs related to diagnosis, assessment, and evaluation in preschool through second grade.

3.1.1 Needs in Preschool

The director of science and technology in the Department for Pre-School Education is in charge of inspection of math education in preschools and in-service training for preschool teachers. Dr. Anat Sela (the national inspector) and Sarita Spokoiny (the national instructor) explained that the preschool math curriculum is based on a sociocultural paradigm that stresses the importance of including mathematical problem-solving in the daily activity in the preschool in order to prepare children to function well in society and master their environment. Preschool mathematics takes the form of structured games and routine functional activities. The teachers are instructed to carry out two main diagnostic and evaluation activities:

1. Reflecting (e.g., “What did I plan to do?” “What actually happened to me and the child?” “What did I do with what happened?”). The teachers are trained to reflect in workshops led by instructors and inspectors.
2. Gathering data in the course of preschool activity through observation of structured play and conversations, and analyzing the data

The data are gathered, collated, and analyzed with the Mabatim tool developed by the Department for Pre-School Education. The teachers have an observation sheet for writing down the goals of the observation and its context (e.g., games, materials, papers written by the children). Dr. Sela stressed that the inspectors' educational conception favors internal evaluation by preschool teachers, and that it is important that every tool developed for this purpose take advantage of unstructured situations in the preschool, be used solely by the teacher, help the children make progress, and not involve labeling.

The following are the diagnostic, assessment, and evaluation needs for preschool mathematics as stated by Dr. Sela:

1. Mabatim-compatible observation sheets based on the revised curriculum to help preschool teachers carry out observations during free play. The observation sheets should include a list of variables to be observed; the teacher can choose those variables that are relevant to the goal of the specific observation. These sheets will help the teacher map out each child's level of functioning and mastery of various skills and plan the teaching.
2. Discourse-analysis tools for evaluating comprehension of mathematical subject matter
3. Tools for standard analysis of artifacts of specific tasks (as opposed to worksheets, which mainly describe educational inputs). Dr. Sela noted that tools for analyzing the children's work and drawing c

onclusions are lacking.

3.1.2 Needs in Elementary School: Grades 1–3

Even though it was decided that the Investigating Panel would focus on diagnosis, assessment, and evaluation in preschool through second grade only, its list of needs pertaining to mathematics relates to third-graders as well, as is standard in the Western world. Dr. Chana Perl (the chief inspector) and Tammy Giron (the national instructor) explained that the tests used today to evaluate mathematical functioning and achievement include tools of various sorts: class tests summing up a topic, internal or external summative tests such as the chief inspector's tests and the GEMS tests, tests to identify sources of difficulty with a particular topic, formative evaluation tools, didactic diagnoses for identifying disparities in a field of content, and tests to diagnose deficits in quantitative matters. Tammy Giron noted the large number of tests used by the elementary education system, most of them developed locally, that have not been standardized and for which psychometric reliability and validity have not been examined. She added that the Israeli elementary-school math curriculum explicitly recommends using routine evaluation tools—tests, quizzes, conversations, observations of pupils' working methods, monitoring of how they cope

with complex tasks, and assignments involving reflective writing. It is recommended that various kinds of evaluation tasks be used: short tasks to evaluate mastery of skills; short tasks to evaluate formal and informal understanding of concepts and the ability to use problem-solving strategies; and complex tasks to evaluate comprehension of concepts, the ability to use them, and the ability to combine and link different areas of mathematics.

The following are the diagnostic, assessment, and evaluation needs in mathematics in grades 1–3 as stated by Dr. Chana Perl and Tammy Giron:

1. Tests are needed to identify sources of difficulty in learning a particular topic (basic knowledge): For first grade a test is needed to diagnose mastery of basic material and basic skills. The test should have several levels of difficulty and entail a variety of evaluation methods (e.g., oral and written tasks, combinations of skills) and should make it possible to identify sources of difficulty in the very first stages of learning. The inspector recommends administering the test at the beginning of second grade. The national instructor proposed a possible model of subject matter and skills, mastery of which should be evaluated in the first-grade test.

Topic	Skills
Number recognition (up to X)	Counting, continuing counting, counting objects, counting part of a group, reading and writing, sequencing, comparing, estimating
Mathematical symbols	Reading, writing, direction of writing math problems
Addition and subtraction: storing quantities in memory, various implications of mathematical operations, mastery of facts up to X, relationship between operations	Writing math problems, recognizing the whole and its parts and being able to distinguish between them, representing a problem with accessories, converting a situation into a math problem, engaging in mental computation

2. Tools are needed for didactic diagnosis of disparities and deficits in the field of content. For grades 2–3, modular tools are needed to evaluate mastery of topics and skills included in the curriculum. These tools should make it possible to identify sources of difficulty in learning various aspects of a particular topic by means of a variety of tasks (e.g., problem-solving, routine and non-routine word problems, inquiry questions, and numeracy items) and in a way that tests both knowledge of the topic and knowledge of other topics. It is important that the tools include tasks requiring high-level thinking (unlike the existing tools, which mainly involve basic-level tasks that require the use of standard algorithms and do not require linking the various topics studied).

3. A tool is needed to evaluate the quality of mathematics instruction so as to help demonstrate math classes to fellow teachers, instructors, principals, and inspectors who observe them.
4. Tests are needed to diagnose arithmetic disabilities. In this context, Dr. Miriam Ben-Yehuda stressed that diagnosis of children in preschool through second grade must be dynamic and must focus mainly on describing existing abilities and potential for improvement. The diagnosis should be based primarily on conversations with the child and on observation. Dr. Ben-Yehuda explained that difficulties observed at a young age should be treated as developmental delays and not disabilities due to the highly dynamic nature of development in young children. She presented a group of tools used to diagnose mathematical functioning (see the description of the present situation below) and stressed that most of the tools have not been standardized and that there is a shortage of high-quality standardized tools.
5. Tests are needed to identify students who are gifted in mathematics.

3.2 The Present Situation: Survey of Evaluation Tools for Mathematics Achievement and Skills Tests Developed in Israel

The Mabatim program developed by the Department for Pre-School Education in the Ministry of Education evaluates a limited sample of arithmetic skills (quantitative and spatial concepts as part of the focus on cognitive activity) by means of observation of preschool children by their teacher. The child's observed success is described on a four-point scale (always/almost always/almost never/never).

Standardized Diagnostic Tools in Hebrew for Children Aged 3–8

1. Prof. David Tzuriel's diagnostic method (Tzuriel, 1998). The method is based on a set of tools to test mastery of a series of skills relevant to arithmetic (e.g., sequencing). No information on norms is available.
2. The Elul test kit: a battery of tests for grades 1–3, with Israeli norms, to identify difficulties in mathematical functioning. So far the Investigating Panel has not been able to examine the content of the tests or data attesting to their quality.
3. The Key Math test, which is used mainly for research. According to experts, the tool is not sufficiently gradated (there are excessively sharp gaps between its requirements), and therefore researchers and diagnosticians often use only some of the tasks included in it. The test lacks Israeli norms.
4. The arithmetic subtest of the Kaufman test (K-ABC). This test of general mathematical functioning includes questions on diverse subject matter that reflect mastery of various skills.

The test was developed as part of a battery of tests for measuring IQ. It can serve only as a rough tool for identifying children with difficulties.

5. The Wechsler arithmetic subtest. The test was developed as part of a battery of tests used to measure IQ. This test too, in the WPPSI version for early childhood (and the WISC-R version for ages 5–16), examines general mathematical functioning and therefore can serve only as a rough tool for identifying children with difficulties.

Non-Standardized Diagnostic Tools in Hebrew for Children Aged 3–8

1. Diagnostic tools developed in Israel (e.g., tests by Rivka Kidron and Dassi Segal). We have no information about norms.
2. Dr. Miriam Ben-Yehuda's diagnostic tool, which includes a set of tasks for examining mastery of mathematical subject matter and basic thinking skills. The tool is based mainly on analysis of a conversation with the child and analysis of errors. We have no information about norms.

Standardized Diagnostic Tools That Have Not Been Translated into Hebrew

1. Woodcock-Johnson test: An intelligence test for ages 2–8 that includes a battery of achievement tests to measure mastery of arithmetic operations, ability to solve arithmetic problems, and familiarity with and understanding of quantitative concepts
2. Brigance: A test administered individually to children aged 2–5
3. Diagnostic Inventory of Basic Skills and Learning Styles in Mathematics
4. Stanford Diagnostic Mathematics Test, 9th edition (1996)
5. Wide Range Achievement Test (WRAT revised)
6. Vernon Miller Test (1986)
7. Basic Number Screening Test (Gillham & Hesse, 1987)
8. Mathematics 8–12, NFER (Brighthouse, Godber, & Patilla, 1984)
9. Test of Early Mathematics Ability (Ginsburg & Baroody, 1990)

It is important to note that we have no tests in Arabic for diagnosing, assessing, and evaluating mathematical learning in early childhood.

3.3 Recommendations

The main questions that the Investigating Panel addressed were as follows: (1) What should be evaluated? (2) Why should it be evaluated? (3) Who should be evaluated? (4) How should the evaluation be carried out? (5) At what age should mathematical knowledge and mastery of various mathematical skills be evaluated?

As stated above, the goals of teaching and the mathematical subject matter covered at each age in Israel are decided on by the subject committee, and its decision is binding on the entire education system. Therefore, the mathematics sub-team focused mainly on formulating recommendations regarding how to evaluate the learning of the material and achievement of the curricular goals. Nevertheless, it should be noted that the education system does not have research-based knowledge of the typical functioning of Israeli children in preschool through second grade in the two main mathematical realms relevant to these ages: numbers and operations, and geometry and measurements.

3.3.1 General Recommendations

1. We recommend conducting a comprehensive study to obtain information on normative mathematical knowledge and mastery of skills among children in preschool through third grade in Israel. The study should address the main mathematical subject matter learned in these age groups. Such a study is an essential first stage in determining what subject matter and mathematical skills should be evaluated at these ages. We recommend that the study address various types of mathematical knowledge and mathematical skills described in the literature, such as formal, algorithmic, and intuitive knowledge (Fischbein, 1993); and conceptual understanding, procedural fluency, strategic competence (the ability to formulate, represent, and solve mathematical problems), and adaptive reasoning (Kilpatrick, Swafford, & Findell, 2001). A wide variety of tasks should be used to examine the various types of knowledge.
2. We endorse the recommendations presented in the chapter on evaluation in the math curriculum for grades 1–6 (Ministry of Education, Culture and Sport, 2006), which state that evaluations should address thinking processes and methods of problem-solving and not just the final results; should monitor pupils' actions, reasoning, and thoughts while they are engaged in mathematical activity; should not limit evaluations to merely giving grades; should use diverse methods; and should look at cognitive issues related to content (understanding concepts and tasks, being able to cope with problems, developing proficiencies, learning communication skills, displaying originality).
3. We recommend diagnosing suspected quantitative deficits during evaluations of normative mathematical functioning (knowledge, content, and skills). Difficulties in mathematical functioning at a young age are liable to stem from a problem with the quantitative cognitive functioning that underlies quantitative thinking. When functioning is particularly poor and does not improve with appropriate instruction and practice, mastery of cognitive functioning should be evaluated in order to set up an appropriate intervention program as early as possible.

Therefore we recommend developing a standardized diagnostic tool to evaluate quantitative cognitive functioning in order to identify mathematical disabilities (dyscalculia). The tool should have countrywide norms and should be used to diagnose young children as early as possible, although evaluators should be aware of and sensitive to the rapid changes that occur at these ages.

4. We recommend compiling an inventory of the major characteristics of outstanding mathematical ability at a young age and developing suitable assessment tools for identifying children who are gifted and talented in mathematics.
5. We recommend assessing—on the individual and systemic levels—pupils’ attitudes toward math and their sense of their own ability to learn the subject. In recent years there has been increased awareness of the need to evaluate pupils’ attitudes toward math, in the understanding that some of the barriers to learning stem from negative attitudes and loss of confidence in one’s ability to learn math. This aspect is included among the goals of the curriculum for preschool (“fostering a positive attitude toward math” [Ministry of Education, Culture and Sport, 1994/95]) and grades 1–6 (“preventing failure and a sense of failure and getting pupils to like the subject” [Ministry of Education, Culture and Sport, 2006b]). The definition of a mathematically proficient child, according to Kirkpatrick et al. (2001), includes not just knowledge of content and processes, but also a “productive disposition—habitual inclination to see mathematics as sensible, useful, and worthwhile, coupled with a belief in diligence and one’s own efficacy” (p. 5). The chapter on evaluation in the math curriculum for grades 1–6 also states that it is important for every evaluation of math courses to address emotional and social factors such as self-confidence, ability to work as part of a team, a positive attitude toward the subject, diligence, and a drive for independence (Ministry of Education, Culture and Sport, 2006b). One way of evaluating attitudes would be to develop a tool for evaluating mathematical functioning that looks at pupils’ attitudes toward the subject and helps identify attitudes that are liable to detract from learning math.

3.3.2 Recommendations for Preschool Education

The preschool curriculum framework (Ministry of Education, Culture, and Sport, 1994/95) lists eight goals related to the development of proficiencies (observation, investigation, and reasoning) and skills (counting objects, measuring, and comparing) and cultivation of positive attitudes towards math (e.g., curiosity, a desire to investigate and experiment with math, and achievement of confidence and enjoyment from math). It describes a developmental sequence for acquisition of mathematical concepts but does not specify standards for normative performance at the ages in

question. A new preschool math curriculum is about to be published, but it was not available to the Investigating Panel. Therefore our recommendations are as follows:

1. Evaluation of mathematical functioning among preschool children, which is currently conducted mainly by the preschool teacher with Mabatim, is fairly limited. In view of the widespread use of Mabatim and the extensive work put into developing, disseminating, and implementing this tool, we recommend using it as the basis for a more fundamental and comprehensive tool that will improve evaluation by testing a wider range of skills, developing standard observation sheets on which to document performance based on predetermined criteria, and establishing a follow-up and monitoring system for orderly gathering of data on the functioning of preschool children. We suggest that the development of this tool be accompanied by systematic research into the way it is used and its outcomes (levels of performance).
2. At this stage we recommend developing structured, standardized tools for evaluating mathematical functioning (knowledge of content, knowledge of processes, and proficiencies) of preschool children in accordance with the curriculum. Development of the tools should be based on mathematical analysis of the concepts learned in preschool and on the use of practical and research-based knowledge about child development and the normative abilities of children at each age. These tools are essential for purposes of individual and systemic diagnosis, assessment, and evaluation.

3.3.3 Recommendations for Elementary School: Grades 1–3

The recently published math curriculum for grades 1–6 (Ministry of Education, Culture, and Sport, 2006b) focuses on two main fields of content: (1) numbers and operations (including data analysis); (2) geometry and measurements. With respect to numbers and operations, the emphasis in grades 1–3 is on recognition of natural numbers and operations involving natural numbers. The curriculum states that “the new aspect of these topics [numbers and operations in grades 1–3] is the emphasis on insight and investigation.... Pupils can develop nonstandard algorithms, and this will reinforce their understanding of operations involving numbers” [p. 5]). In terms of geometry and measurements, the emphasis is on developing a visual conception in two and three dimensions and on “developing methods of thinking such as deductive reasoning, which is characteristic of math as a whole, coming up with hypotheses and testing them, generalizing, articulating reasoning, and making inferences” (ibid., 8). The curriculum describes in detail the knowledge required with respect to numbers, arithmetic operations, geometry, measurements, and data analysis for each

grade. The last page of the curriculum for each grade sums up the key elements that each pupil should know at the end of that grade.

As presented by the inspectors, the evaluation needs for grades 1–3 are many and diverse. A scale of priorities should be drawn up regarding the development of diagnostic, assessment, and evaluation tools. There are currently a few tools that may be suitable (e.g., Elul), but without access to them it is hard to assess their present suitability. In order to meet the needs of the system, the existing tools should first be surveyed and assessed professionally, with the focus on the need for tools intended for formative evaluation of achievement, evaluation of knowledge and skills, and diagnosis of difficulties in specific areas.

Chapter 3

Summary and Recommendations Regarding the Issues Examined

The Investigating Panel's recommendations are derived from the present situation in Israel regarding each issue examined and a great deal of research information on the various aspects of development and fields of content. The recommendations were tailored to the needs of the Israeli education system.

We believe that standardized tools should be developed for diagnosing, assessing, and evaluating children's level of functioning in the following areas: socio-emotional development, language, linguistic literacy, and mathematics. In addition, new tools are needed for evaluating the climate in the preschool, class, or school. The evaluation needs in preschool through third grade are as follows: (1) tools for a formative evaluation of achievement; (2) tools for evaluating knowledge and skills; (3) tools for evaluating difficulties in specific areas. Thus, practical, action-oriented tools are needed. Developing and implementing such tools will require extensive cooperation among agencies within and outside the education system, as well as a commitment to initiating intervention programs to promote socio-emotional adjustment, language and linguistic literacy abilities, and the learning of math in these age groups.

1. Recommendations for Further Research and Development

Evaluating the functioning of children aged 3–8 in the education system encompasses two separate stages, each with its own unique goals.

The first stage is screening. This involves (1) surveying the unique styles and needs of children in the preschool or class in order to tailor the teacher's work plan to them; (2) identifying children with special needs, children at risk with developmental delays, and gifted and talented children, in order to refer them for a more in-depth evaluation so that an individualized intervention plan can be designed ; (3) surveying the condition and needs of children on various levels of the education system so that the Ministry can plan and make decisions. These goals led to a recommendation to follow all children entering the education system from preschool on and to monitor their development using standardized tools.

The second stage is diagnostic evaluation of those children identified in the screening stage as having difficulties. This is an inclusive, detailed evaluation of special needs related to various

aspects of functioning dealt with by the committee (emotional, behavioral, social, and cognitive) and various fields of study (language, linguistic literacy, and mathematics). The purposes of this evaluation are (1) to gain a deeper understanding of the child and his/her environment as an integral unit in order to plan an appropriate intervention program for prevention and treatment; (2) to create a foundation of mutual understanding with each child's environment and to become familiar with the child's unique needs in order to reach agreement on goals and working methods; (3) to survey the condition and needs of special-needs children on the various levels of the education system so that the Ministry staff can make well-informed decisions.

A valid, reliable standardized tool with countrywide norms should be developed to enable the education system to identify children with special needs, map out their specific needs, and develop intervention programs that meet these needs as early as possible. The tool should be a developmental survey and should not label the children with definitive diagnoses. It should evaluate their performance, their cognitive, emotional, behavioral, and social functioning, and their knowledge and skills with respect to language, linguistic literacy, and mathematics on a normative spectrum. It should also address aspects of the development of knowledge and learning in early childhood that are outside the purview of the Investigating Panel. The tool should be designed with a multicultural society and the need to evaluate special populations in mind. Such a tool can also help with mapping geographical regions with special needs and evaluating the progress of the education system over time.

1.1 Guiding Principles for the Development of Screening Tools

We recommend learning from the experience of developing Mabatim, implementing the positive principles that characterized it, and correcting the flaws and deficiencies. Drawing conclusions from previous developmental work may be helpful to the Department for Pre-School Education and other departments that need a screening tool suitable for ages 3–8.

We recommend adopting the principles that characterized the development of the Mabatim program:

1. Development of the tool was led by Education Ministry personnel who were thoroughly familiar with the work of the homeroom teacher, points of encounter between the homeroom teacher and the child, and the homeroom teacher's ability to adjust his/her work to the various needs of the children in the class.
2. The tool uses observation-based, practical language that does not label the children. For screening purposes, we prefer to use terminology based on the tradition of community work, on a survey of development in various fields of knowledge, and on follow-up after this survey.

3. In addition to the booklet that goes along with the observation, an observation manual for teachers was written and a program for training homeroom teachers in methods of observing various aspects of functioning was devised. The training program enables homeroom teachers to develop and enhance their ability to conduct detailed observations of their pupils' functioning.
4. Another program developed is Ma'agan, which instructs teachers in how to intervene and adapt their work to the needs of the children as described in the goals of the observation. In other words, the homeroom teacher's investment in using the tool enhances his/her work with the children and their parents.
5. The tool was developed concurrently for several population groups, including Arabs, the ultra-Orthodox, and immigrants. The observation tool and the intervention tool were written in both Hebrew and Arabic.
6. In addition to the time put into developing the observation tool and the intervention program, a budget was allocated for training homeroom teachers, distributing the tool, and introducing it in the schools. Without a suitable budget, the time, effort, and money invested in developing unused tools would be wasted.

The drawbacks of the Mabatim program were also noted, and other working principles were articulated that can be implemented in Mabatim and in the development of new tools.

7. The Ministry staff that spearheads the development of the tools should include RAMA personnel who are charged with developing valid, reliable tools and are familiar with the necessary procedures for determining validity and reliability.
8. Academics should be involved in developing the tools, both by working with the Ministry staff and by carrying out field research on which the development and validation of the tools can be based.
9. In view of the tendency in screening to emphasize the cognitive dimension and de-emphasize the socio-emotional dimension, and in the wake of the research findings and practical experience indicating the centrality of the socio-emotional dimension in children's adjustment to the education system, it is important for the development team to include experts in socio-emotional development as well as in language, linguistic literacy, and mathematics.
10. Evaluations of schoolchildren (not preschoolers) can use a tool that is filled out by the children themselves. Tools of this sort are used abroad to evaluate children starting in second grade, and we can learn from them, translate them, and adapt them to our pupils.

1.2 Validation of Screening Tools and Development of Norms

We suggest investing judiciously in validating screening tools and developing suitable norms.

However, because this involves a substantial investment of resources, we recommend doing it only if the following conditions exist:

1. Decision-makers in the Ministry of Education demonstrate clear motivation to do the job and are contending properly with practical concerns about labeling and irrelevant diagnoses (diagnoses with no practical consequences that remain in a drawer).
2. There is a multi-subject team that encompasses all the personnel needed to accomplish the task, including academics and others involved in the development of evaluation tools.
3. Decision-makers in the relevant units of the Ministry of Education are prepared to invest in researching, developing, evaluating, distributing, and introducing intervention programs that apply the findings of the screening work. Even if the development and implementation of such programs are merely preliminary, we recommend investing in the programs while also developing the screening tools. Because many intervention programs devised by different agencies have not been evaluated by means of orderly research and according to the present standard criteria for evaluation of programs, and because these programs require complex collaboration among various departments in the Ministry of Education, methodical work by the Ministry staff is needed to systematize the activities. For this purpose as well, Ministry personnel can consult academics and others in order to design the programs and ways of evaluating them. At first it may be necessary to look into the possibility of having the national Ministry staff draw up a theoretical flowchart of methods of intervention to be used by the various services in Israel, with interdepartmental cooperation among the Psychological Counseling Service, the Department for Pre-School Education, the Department for Elementary School Education, the Department for Special Education, the Department for Talented and Gifted Students, the Department for Arab Education, the Department of Education and Welfare Services, the Immigrant Absorption Unit, and the chief inspectors of Hebrew and Arabic for native speakers in the Pedagogical Secretariat. The guiding principle should be that the more we invest in improving screening tools, the more we enhance the intervention programs. For instance, Mabatim is primarily a tool with which preschool teachers can evaluate children, but if the evaluation needs are defined more clearly and addressed in an optimal manner, the teacher's individual and group work with the children and their parents will be furthered.
4. The efforts to ensure validity and reliability and to create norms, as well as the work on designing intervention programs, must take into account various population groups, including Arabs, the ultra-Orthodox, and immigrants, from the very beginning. It should be kept in mind

that criteria for cultural adaptation of screening, assessment, and evaluation tools have been disseminated by the International Testing Committee (ITC), and there is professional knowledge available that can help increase the cultural sensitivity of the people carrying out the evaluations and interventions.

5. As stated above, there is no need to invent screening tools from scratch; one can rely on the experience gained in developing Mabatim. It is also worth becoming familiar with tools that have been developed and applied abroad, such as SDQ, which can be used in Hebrew without paying royalties to the authors.

1.3 Diagnostic Evaluation

At the end of compulsory preschool, a decision must be made as to whether each child is ready for first grade or should spend another year in preschool. These decisions are currently based on a diagnosis of first-grade readiness of those children whose readiness is in question, according to the criteria specified in *Director-General's Bulletin 9* for holding children back in preschool (Education Ministry, Culture and Sport, 2004/05). However, since this *Director-General's Bulletin* came out, new research-based information has accumulated about indicators that are predictive of the effect of holding children back for another year of preschool. We recommend developing a decision-supporting computer system (similar to the one developed by the Youth Probation Service), backed by ongoing research and based on valid and reliable structured diagnostic tools, that revises the criteria every year. We are not talking about making decisions on the basis of a computerized report but about systematically gathering and processing data so that informed decisions can be made.

It should be stressed that everything stated in other parts of this report applies to children's readiness for school as well. Identifying children who are having difficulty in preschool and then diagnosing and treating children who need special assistance are a major part of the service that must exist so that children can adjust to preschool, make a healthy transition from preschool to school, and adjust to school.

2. Guiding Principles for Screening and Diagnosis

2.1 Parental Involvement

The decision to conduct a diagnostic evaluation of specific children who are either having difficulty or are gifted and talented should be made in consultation with the preschool teacher, a psychologist, a guidance counselor, and the parents. Children should be referred to additional professionals such as psychiatrists and pediatric neurologists only after such consultation. Due to the present structural

separation between preschool and elementary school, the Investigating Panel recommends that every diagnosis, evaluation, or special intervention in the preschool be brought to the knowledge of the relevant people in the elementary school that the child will be attending, in order to maintain continuity in the work with each child and the child's family even after the child enters first grade. Orderly, systematic communication of this information will ensure that every special-needs child entering first grade receives continuous, appropriate treatment, as well as whatever extra educational investment is needed. Naturally, information about a child can only be communicated after the parents have been informed and given their permission.

We also recommend involving the children's parents in all information-gathering processes for screening and, of course, diagnostic evaluation purposes and informing them that the educators, including the psychologists in the preschools and schools, are following the children's development while they are in the education system (just as well-baby clinics follow the development of infants and young children). Establishing a work alliance, collaboration, and positive interrelations with parents will make it possible to ask the parents to fill out a non-threatening, non-labeling questionnaire with background information about the family and the child. Obviously, a system should be set up to train preschool teachers, schoolteachers, psychologists, and guidance counselors to conduct interactions of this sort with parents. The training of educators and establishment of mechanisms for bringing about parental involvement will be especially necessary if a decision is made to change the current diagnostic tradition and switch to widespread use of more valid, more reliable, and more systematic structured tools for diagnosis, assessment, and evaluation.

Involving the parents should include providing feedback as an integral part of the screening and evaluation of their child, whether the child is found to have special needs or not. It would be best for the psychologists and guidance counselors to be present when teachers give feedback to parents, but given the personnel shortage in educational institutions, a more practical recommendation would be to invite a psychologist to the feedback meeting only in the case of a child at risk or if the relationship between the teacher and parents is problematic.

2.2 Insistence on Ethical Principles and Confidentiality

All information gathered about the children and their families should be kept confidential, and should be passed on to others only if the parents request this and sign a waiver-of-confidentiality form. However, there should be a procedure for passing on information without identifying details for the purpose of an overall developmental survey of the children's condition and needs. Such a procedure would make it possible to keep the personal information secret while also using it for

policymaking and decision-making regarding allocation of resources, planning of educational interventions, and evaluation of the interventions.

2.3 Evaluation of the Climate in the Class and Educational Institution and of Teacher-Pupil Relations

Any effort to diagnose, assess, and evaluate children in general and their socio-emotional development in particular must address their family context and home environment, as stated above, but also the children's educational environment and the educational climate in the preschool or school. We are referring here to interpersonal relations between teachers and children and actions taken to meet the children's developmental needs. These aspects affect the children's adjustment to the preschool or school in the present and their future in the education system. Moreover, evaluations of individual children must take into consideration the environment in which the children function and are being tested. Structured procedures and tools are available abroad for systemic evaluation of the educational climate by psychologists who visit preschools and classes, the instructors who train the homeroom teachers, and the inspectors of their work. Without structured, validated tools, evaluation is liable to be merely a matter of impressions—not sufficiently detailed, not valid, and not reliable. We recommend that decision-makers in the Ministry of Education consider adopting appropriate tools developed abroad and/or creating original tools through cooperation among various Ministry agencies (e.g., the Psychological Counseling Service and the inspection system) and experts from academia. Because the climate of the educational environment has a decisive impact on young children's development, it should be monitored and intervention programs should be developed that address this systemic level as well..

3. Fundamental Methodological Recommendations

The Investigating Panel drew up several fundamental methodological recommendations concerning ways of developing and designing tools for diagnosis, assessment, and evaluation. The recommendations below concern evaluation of the quality, validity, reliability, and fairness of the tools, as well as norms and standards. Below we use *test* as a general, broad term that encompasses the entire range of diagnostic, assessment, and evaluation tools reviewed by the Investigating Panel.

The use of standardized scores in this report is reserved solely for tools meant to evaluate basic cognitive and behavioral functions. These evaluations are intended mainly to identify children who are functioning below the national norm in these respects. Most of these functions are neurologically based. They are not learned directly in the education system and therefore no criteria for performance have been set for them. In contrast, performance on tools meant to evaluate

learning should be examined relative to criteria or standards derived from the curriculum or set in advance by experts.

The psychometric literature suggests a variety of strategies and methods for examining the quality of diagnostic, assessment, and evaluation tools. When choosing a tool, one should keep in mind the assessment needs and the use that will be made of the data (American Educational Research Association, American Psychological Association, and National Council of Measurement in Education, 1999).

3.1 Development and Design of Tests and Evaluation of Their Quality

This process must be based on stable, up-to-date theoretical knowledge and research and must state the following clearly and explicitly:

- The purpose of the test
- The content and/or skills tested
- The rationale for the test items
- The target population
- Acceptable uses of the test results
- A method of analyzing the data gathered in the test: criteria for scoring each test item; scale(s) for reporting results; a procedure for calculating the indices summing up the test results; and guidelines for interpreting the results. When several versions of a test are administered simultaneously or on different dates, the different versions should be calibrated so that the scores can be compared. The reasoning behind the scoring criteria, the reporting scales, the procedures for calculating indices, the methods of calibrating the various versions, and the significance of these should be specified.

The quality of the test should be evaluated by means of a pilot test that examines the test items and the distribution of test scores.

3.2 Validity

The purpose of the examination of validity is to present logical arguments and evidence from research showing that the test achieves its stated objectives and is suitable for examining these objectives among the target population(s).

- The relationship between the test items and the content and/or skills examined (content validity) should be established.

- In predictive tests, the relationship between the test results and the criterion that the test is meant to predict (predictive validity) should be established; i.e., statistical evidence should be provided for the relationship between the score and the criterion.
- In group screening and placement tests, evidence of the test’s differential validity should be provided for each group.
- In diagnostic tests, evidence should be provided for the validity of the distinction between a “typical” population and a “clinical” population.
- The impact of practice or direct preparation by test-takers on the validity of the test should be examined.
- The theoretical rationale for each interpretation given to the test results should be specified and any use of these results should be justified.
- In the development or adoption of diagnostic, assessment, and evaluation tools, the ecological impact of the tools should be taken into account. This refers mainly to possible consequences, whether positive or negative, of the use of tests or of decisions derived from the results of such use. Although ecological impact considerations are important in all contexts, they are especially important in the diagnosis of young children. Therefore, when choosing tools to be used with children aged 3–9, one should take into account their impact on the teacher’s work, on changes in emphasis in the curriculum and ways of teaching it, on the danger of labeling children at such a young age, and on erroneous placement of children in educational settings. Good diagnostic, assessment, and evaluation tools are those that empower teachers, contribute to the development and improvement of relevant skills and good teaching, and prevent unnecessary and irreversible labeling of children.

3.3 Examination of Reliability to Minimize Measurement Error

A good diagnostic evaluation is based on reliable tools that meet the standards of similar diagnostic, assessment, and evaluation tools.

- The reliability of the overall test score and the score on each individual scale should be examined. Reliability checks generally look at internal reliability and test-retest reliability. When possible, more than one type of reliability should be checked and the type of reliability coefficient calculated should be specified.
- When test scores are based on human evaluators, at least three evaluators should be used and the reliability should be calculated by comparison of their judgments.
- The measurement error for each scale should be computed and reported.

- The reliability of the test and the measurement error should be computed for each population subgroup whose traits differ from those of the majority.
- Above we recommended examining the test-retest reliability of every tool used for diagnosis, assessment, and evaluation. In this context, however, it is important to stress a serious problem with administering tests to young children: the level of test-retest reliability. Test-retest reliability is important in evaluating the stability of people's traits over time, especially when decisions regarding a student's future education assume that the trait is stable and unchanging. On the other hand, it is important to note that test-retest reliability is less relevant to evaluating achievements of young children. It is hard to develop a tool on which the same child would receive the exact same score on a retest administered a few days or weeks later. The rate of maturation of cognitive functioning varies from child to child, so the scores are not stable and test-retest reliability is low. Consequently, it is very important to examine the stability of all the types of functioning tested by means of test-retest reliability at short time intervals among children whose functioning is just below or just above a cutoff point. This retesting can be limited to evaluation tools with low test-retest reliability and children whose functioning is close to the cutoff point.

3.4 Fairness

- It is important to ascertain that the test items are not biased for or against specific groups of test-takers.
- Test items that are liable to be prejudicial to test-takers due to differences of gender, religion, race, and the like should be identified and disqualified.
- It is important to make sure that test-takers' linguistic ability does not detract from the validity of the test unless it is relevant to the objectives of the test.

3.5 Norms and Standards

- Norms should be set based on a representative sample of the target population.
- On criterion-referenced tests, standards of performance should be set through the use of procedures that are customary in the professional literature.
- The stability of the norms or standards should be reviewed periodically and they should be adapted to the needs.

4. Standards for Distributing and Introducing Tests

The following information and materials should be provided to potential users of the tests:

1. A general users' manual, which should include the following information:
 - The purpose of the test, the target population, and acceptable uses of the results
 - The rationale for the test, including theoretical and research background and a detailed description of the development process
 - Information about the quality of the test: validity, reliability, and fairness
 - Norms or standards for performance (cutoff points and the significance of levels of performance indicated by them)
 - Information about the use of the test among different groups of test-takers
 - Information that will reduce the likelihood of a misunderstanding or incorrect use of the test results, including a warning about inaccurate interpretation of the results and unacceptable uses of the test
 - The professional qualifications needed to administer the test and evaluate performance on it
2. A kit containing all materials needed to administer the test, including standard response sheets if necessary
3. Detailed instructions for administering the test, including information about conditions for administering it, such as instructions to test-takers that are not written on the test itself, reference materials that may be used, modifications for special population groups, and the time allotted
4. Instructions for evaluating performance: rules and/or principles for evaluating performance on each item and instructions for producing scale scores and overall scores.

General Requirements

Every diagnostic, assessment, and evaluation tool developed must include the following components:

1. A standardization mechanism for checking quality, a seal of quality, documentation and research, supervision and periodic evaluation (to determine the ecological impact), processes of change and improvement
2. A system for training people to administer the test and to analyze, interpret, and use the results
3. Standards for the desired scope of evaluation activity: quantity, frequency, integration in regular work, and balance among parts of the evaluation

4. A firm decision as to the dates of the evaluation during the year
5. A specific explanation for preschool teachers as to the outcomes of the evaluation from a therapeutic standpoint

5. Diagnosis, Assessment, and Evaluation in a Multicultural Society

(based on the conclusions of the Joint Committee on Testing Practices [JCTP], 2004)

Four groups in the multicultural Israeli society require special attention in the planning and development of diagnostic, assessment, and evaluation tools, use of the tools, scoring, and interpretation of the results: immigrants whose native language is not Hebrew (from the former Soviet Union, Ethiopia, etc.), Arabs, the ultra-Orthodox, and individuals with special needs, including pupils with physical and/or learning disabilities and gifted and talented pupils. First, basic research must be conducted to gather information about the unique features of these population groups in this context and to understand their ramifications. For instance, a study should be done to clarify the differences between the Arabic language spoken at home and the language used in school, or between written Biblical Hebrew and the spoken language used in the haredi sector. Presumably, the issues evaluated in these sectors are similar but the methods, tools, and tests used to evaluate pupils in each of these groups must have unique features.

In order to ensure the validity, reliability, fairness, and usefulness of tools designed for diagnosis, assessment, and evaluation of these population groups, the following conditions must be met:

1. A task force should be established and standardized processes instituted to cope with the multifaceted structure of a multicultural society in which some segments of the population do not know the language and culture on which the standard psychological and educational tests are based.
2. Tests meant for members of specific population groups should suit the traits and background of those groups. It is important to focus on research-based and practical classifications of the population groups, avoid labeling, and develop separate indices for the different groups.
3. When norms are used, they should relate to the relevant group. Test-takers should be evaluated relative to the norms developed for their group. In setting the norms, it is important to take variations among groups into account from the very beginning of the development process and not only after the process is completed.
4. When tests have been translated from another language and adapted to the Israeli education system, the methods used to establish the accuracy of the translation and to produce logical and

research-based evidence of the validity and reliability of the conclusions with respect to the Israeli target population should be described.

5. When comparisons are to be made among versions of a test in different languages, evidence should be provided establishing that the different versions are identical.
6. Test results should be interpreted in light of the material that the test examined, norms or comparison groups, other technical evidence, and the benefits to be derived from the test.
7. When scores are computed and published for subgroups, a warning should be attached stating that they may have different implications for the different subgroups.
8. An evaluation system should be created that will minimize threats to the reliability of the test resulting from language differences. The test should be taken in a language in which the test-taker is proficient, unless mastery of another language is part of what is being evaluated.
9. Evaluation tools should not include subject matter or expressions that are liable to offend members of any group.
10. The developers of an evaluation tool should attempt to determine its quality by testing it on samples that allow for analysis by subgroup, and should provide evidence of validity, reliability, fairness, and usefulness with respect to test-takers from different subgroups.
11. Determining the sample size required to ensure the validity of statistical conclusions is a complex issue affected by various factors: the statistical power needed to identify an effect of a given size, interpersonal differences with respect to the phenomenon being measured, and the sampling method. Therefore, it is impossible to formulate a uniform recommendation; the calculation should be based on the circumstances of the case and the statistical analysis desired. For instance, the sample size required to estimate the difference between the averages of two groups will be much smaller than that needed to determine norms for assessment tools among a particular population. Some say that 30 test-takers is the minimum sample size, based on the simple case of a random sample and on the properties of a normal sampling distribution. Others claim that the minimum sample necessary for the experimental use of a structured evaluation tool (e.g., a test) is about 200 pupils. Flexible evaluation tools are customarily used on smaller samples. The standard sample for coming up with performance norms is 500 pupils, provided that it is a representative national sample. In the case of an analysis of different groups, the sample size is derived in part from each group's share of the population, the quality of the sample, and the purpose of the test. This recommendation may be appropriate in some cases but should not be seen as a rule of thumb. We recommend computing the sample size required in each individual case in accordance with the procedures used for this purpose (see, e.g., Kraemer & Thiemann, 1987).

12. Users of an evaluation tool should evaluate the available evidence as to performance of the different subgroups and should determine, insofar as they can, what differences in performance may stem from factors unrelated to the content and/or skill being assessed.
13. The developers of an evaluation tool should provide information that will enable users to interpret the results among different groups correctly. This information should address traits of test-takers in the comparison groups and all factors that may influence interpretation of the results.
14. Opportunities should be created for information exchanges between experts in psychometrics and in the development of diagnostic, assessment, and evaluation tools and experts in the study of social and cultural differences between population groups.

For students with physical or learning disabilities, the following special standards must be ensured:

1. Test materials and procedures appropriate for test-takers with special needs who require special accommodations should be provided and documented.
2. Research methods should be used to determine how much time persons with disabilities should be given to take a modified test.
3. When extra time is allotted for a test, the impact of fatigue on the test results should be considered.
4. Interpretation of the results of tests modified for people with disabilities should take into consideration changes made in the course of the modification.
5. Developers of an individualized test for people with disabilities, the people administering the test, and users should take steps to ensure that the conclusions drawn from the test results accurately reflect the trait measured and not disabilities or other traits that were not being tested.
6. When an individual with a disability is tested for purposes of diagnosis and intervention, the test results should not be used as the sole indicator of the individual's functioning. As many sources of information as possible should be used.

For gifted and talented pupils as well, special standards should be ensured:

1. Whether an individual belongs to the group of gifted and talented pupils should be determined in a comprehensive, well-planned process.
2. Tools for determining whether someone belongs to the group of gifted and talented pupils should assess diverse abilities, skills, and strengths and should enable the individual to demonstrate them.
3. A method should be developed for evaluating the strengths and needs of each gifted or talented pupil in order to plan appropriate intervention.

4. All procedures for identifying gifted and talented pupils should be based on the latest theory and research.
5. Written procedures for identifying gifted and talented pupils should address, at the very least, the pupils' consent to be tested and the pupils' persistence, and there should be an option of terminating the test, re-taking it, and appealing the results.

In this context, the issue of the Arabs, who account for about one-fifth of the Israeli population, should be discussed as well. There is a tremendous shortage of diagnostic, assessment, and evaluation tools designed for this sector and of professionals capable of developing and administering them. We recommend planning, designing, and implementing separate programs and diagnostic, assessment, and evaluation tools for the Arabic-speaking population.

Key Questions

1. Does the policy of the Ministry of Education take into account the different groups of pupils in the multicultural Israeli society?
2. If so, does the policy address various aspects of multiculturalism—the pupils' identity, religion, curriculum, language, evaluation tools, staff development, personnel, and parents?
3. Does the Ministry of Education have a task force or group of advisors to deal with the differences between subgroups in Israeli society? Do the members of this group represent the different subgroups?

6. Systemic Evaluation

An inclusive systemic evaluation is based on a series of indices that summarize and organize information on various subjects that is essential to understanding the system as a whole. There are currently two approaches to designing such series of indices: the mechanical engineering approach, based on the CIPP (Context, Input, Process, Product) model, and the social-interpretive approach, based on “massive description” of the interrelations among system components, with emphasis on social and cultural components. The series of indices make it possible to evaluate a policy, examine its implementation, and monitor the use of intervention programs and curricular materials. A uniform series of indices gives decision-makers a systematic, comparative overview and makes possible analysis and the drawing of intelligent conclusions.

Any inclusive systemic evaluation requires, first, a decision as to what subjects it should include. Next, criteria are set for the level of functioning of the system with respect to each of these subjects, and each subject is assigned a series of research-based indices for which data will be

gathered systematically over time (access to learning opportunities, physical infrastructure, quality and quantity of personnel, study materials, etc.).

Developing and designing a series of indices requires a professional and public discussion of policy-making issues: analysis of the information needs of the system, compatibility between the objectives of the system and the information system, identification of various interested parties, assurance that the information system is being used properly, a balance between internal and external evaluations and between formative and summative evaluations, and creation of mechanisms for continuous discourse among the interested parties.

Systemic evaluation requires gathering data from many different sources of information on the various levels of the education system and organizing the information in a way that will ensure continuity among the stages of education. Standards should be set for evaluating the quality of the information gathered and the quality of the mechanisms for gathering, entering, and organizing the information on the various levels of the system. In addition, methods of analyzing the data and using the findings, including the content and structure of reports on the evaluation process and its outcomes, should be decided on.

A systemic information system offers an opportunity for greater involvement by the research community in efforts to improve the education system. Such a system will, on the one hand, receive information from numerous scholars and, on the other hand, make information available to these and other scholars. In order for this combination of activities to be significant and efficient, it is necessary to articulate the main research questions, to design both short- and long-term research agendas, to create a cumulative database that will be available to the scholars, to develop mechanisms for cooperation among educational and research institutions, and to call for proposals for research studies on various relevant topics.

- References
- Achenbach, T. M., & McConaughy, S. H. (2003). The Achenbach system of empirically based assessment. In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children: Personality, behavior, and context* (2nd ed., pp. 406–430). New York: Guilford.
- American Educational Research Association, American Psychological Association, & National Council of Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Beauchaine, T. P., Webster Stratton, C., & Reid, M. J. (2005). Mediators, moderators, and predictors of 1-year outcomes among children treated for early-onset conduct problems: A latent growth curve analysis. *Journal of Consulting and Clinical Psychology, 3* (73), 371–388.
- Ben-Simon, A., Shani, M. and Cohen, Y. (2005). *The Achievement Test for Acquiring Reading and Writing in Second Grades: Summarizing Report*. The Ministry of Education and Sports, and the National Institute for Testing and Evaluation.
- Ben-Simon, A., Shani, M. and Cohen, Y. (2005) *The Achievement Test for Acquiring Reading in First Grades*. The Ministry of Education and Sports, and the National Institute for Testing and Evaluation.
- Butterworth, B. (1999). *The mathematical brain*. Oxford, UK: Macmillan.
- Committee on Assessment and Evaluation in Education. (2005). *The knowledge base for assessment and evaluation in education: A framework for curricula*. Retrieved from http://www.academy.ac.il/data/projects/34/Assessment_and_Evaluation.pdf.
- Deheane, S. (1997). *The number sense*. New York: Oxford University Press.
- Department for Children, Schools and Families. (2006). *National numeracy strategy*. Retrieved October 2007 from <http://www.standards.dfes.gov.uk/primaryframeworks/mathematics/learningobjectives/Years/Foundation/Countcomapreandordernumbers>.
- Fischbein, E. (1993). The interaction between the formal and the algorithmic and the intuitive components in a mathematical activity. In R. Biehler, R. W. Scholz, R. Straser, & B. Winkelmann (Eds.), *Didactics of mathematics as a scientific discipline* (pp. 231–345). Dordrecht, The Netherlands: Reidel.
- Graziano, P. A., Reavisa, R. D., Keane, S. P., & Calkins, S. D. (2007). The role of emotion regulation in children's early academic success. *Journal of School Psychology, 45* (1), 3–19.
- Joint Committee on Testing Practices. (2004). *Code of fair testing practices in education*. Retrieved May 27, 2007 from http://www.alte.org/further_info/sofia/ak_2_1106.pdf.

- Kilpatrick, J., Swafford, J., & Findell, B. (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academy Press.
- Klein, P. S., & Yablon, Y. B. (Eds.). (2007). From research to practice in early childhood education. Jerusalem: Israel Academy of Sciences and Humanities. Retrieved from <http://www.academy.ac.il/data/egeret/86/EgeretArticles/early-childhood-education-hebrew.pdf> (in Hebrew). For the executive summary in English, see <http://www.academy.ac.il/English/hp.asp?file=asp/projects/projects.asp?txt=1||34>.
- Kraemer, H. C., & Thiemann, S. (1987). *How many subjects? Statistical power analysis in research*. Newbury Park, CA: Sage.
- Measelle, J. R., John, O. P., Ablow, J. C., Cowan, P. A., & Cowan, C. P. (2005). Can children provide coherent, stable, and valid self-reports on the big five dimensions? A longitudinal study from ages 5 to 7. *Journal of Personality and Social Psychology*, 89 (1), 90–106.
- Miller, A. L., Fine, S. E., Gouley, K. K., Seifer, R., Dickstein, S., & Shields, A. (2006). Showing and telling about emotions: Interrelations between facets of emotional competence and associations with classroom adjustment in Head Start preschoolers. *Cognition and Emotion*, 20 (8), 1170–1192.
- Ministry of Education, Culture, and Sport. (1994/95). *Framework Curriculum for State and State-Religious, Arab, and Druze Preschools*. Jerusalem: Ma'alot (in Hebrew).
- Ministry of Education, Culture, and Sport. (2003). *Language Education: Hebrew—Language, Literature, and Culture: Curriculum for State and State-Religious Elementary Schools*. Jerusalem: Ma'alot (in Hebrew).
- Ministry of Education, Culture, and Sport. (2004/05). Keeping children in preschool for an extra year. *Director-General's Bulletin*, 9 (1). Retrieved from http://www.education.gov.il/edun_doc/se9ak3_10_11.htm (in Hebrew).
- Ministry of Education, Culture, and Sport. (2006a). *Foundations of reading and writing. Experimental edition of a preschool curriculum*. Retrieved from http://cms2.education.gov.il/EducationCMS/Units/Tochniyot_Limudim/Portal/TochniyotLimudim/KdamYesodi/KrihaKtiva.htm (in Hebrew).
- Ministry of Education, Culture, and Sport. (2006b). *Mathematics curriculum for grades 1–6 in all sectors. Jerusalem: Ma'alot*. Retrieved from http://cms.education.gov.il/EducationCMS/Units/Tochniyot_Limudim/Math_Yesodi/PDF/Mavo.htm (in Hebrew).

- National Council of Teachers of Mathematics. (2006). *Curriculum focal points for prekindergarten through grade 8 mathematics: A quest for coherence*. Retrieved October 28, 2006, from http://www.nctmmedia.org/cfp/front_matter.pdf.
- Pianta, R. C. (1992). *The student teacher relationship scale*. Charlottesville, VA: University of Virginia.
- Pianta, R. C., Steinberg, M., & Rollins, K. (1995). The first two years of school: Teacher-child relationships and deflections in children's classroom adjustment. *Development and Psychopathology*, 7, 295–312.
- Reid, M. J., Webster-Stratton, C., & Hammond, M. (2003). Follow-up of children who received the incredible years intervention for oppositional-defiant disorder: Maintenance and prediction of 2-year outcome. *Behavior Therapy*, 34 (4), 471–491.
- Satlow, E., & Newcom, N. (1998). When is a triangle not a triangle? Young children's developing concepts of geometric shape. *Cognitive Development*, 13, 547–559.
- Shaw, D. S., Dishion, T. J., Supplee, L., Gardner, F., & Arnds, K. (2006). Randomized trial of a family-centered approach to the prevention of early conduct problems: 2-year effects of the family check-up in early childhood. *Journal of Consulting and Clinical Psychology*, 74 (1), 1–9.
- Shaw, D. S., Lacourse, E., & Nagin, D. S. (2005). Developmental trajectories of conduct problems and hyperactivity from ages 2 to 10. *Journal of Child Psychology and Psychiatry*, 46 (9), 931–942.
- Smith, C. L., Calkins, S. D., Keane, S. P., Anastopoulos, A. D., & Shelton, T. L. (2004). Predicting stability and change in toddler behavior problems: Contributions of maternal behavior and child gender. *Developmental Psychology*, 40 (1), 29–42.
- Tzuriel, D. *Cognitive modifiability: Dynamic assessment of learning potential*. Tel Aviv: Sifriyat Poalim (in Hebrew).
- Webster-Stratton, C., & Reid, M. J. (2003). Treating conduct problems and strengthening social and emotional competence in young children: The dina dinosaur treatment program. *Journal of Emotional and Behavioral Disorders*, 11 (3), 130–143.
- Webster-Stratton, C., & Reid, M. J. (2006). Treatment and prevention of conduct problems: Parent training interventions for young children (2–7 years old). In K. McCartney & D. Phillips (Eds.), *Blackwell handbook of early childhood development* (pp. 616–641). Malden, MA: Blackwell.

Webster-Stratton, C., Reid, M. J., & Hammond, M. (2004). Treating children with early-onset conduct problems: Intervention outcomes for parent, child, and teacher training. *Journal of Clinical Child and Adolescent Psychology*, 33 (1), 105–124.

Appendix B

Diagnostic, Assessment, and Evaluation Tools Used in Early Childhood Education: Proposal for an Evolving “Map”

The following is meant to be an evolving table. Revisions should be made as additional information is obtained.

An asterisk (*) indicates a function that the test examines. In the columns pertaining to norms, a blank space denotes insufficient information.

Test	Age	A: 3–6 B: 6–9 A+: 3–9	Education Ministry dept. that uses it	Developer	Year published	Hebrew translation and norm sample (N)	Arabic translation and norm sample (N)	Foreign norms (N)	Aspect of functioning tested	Emotional/behavioral
HATAF	1–3	A	Education and Welfare Services						Home guidance in nurturing toddlers, with emphasis on parent-child interaction and encouragement of inquiry and discovery	
PACT	3–5	A	Education and Welfare Services							
HA’ETGAR	3–6	A	Education and Welfare Services	NCJW Research Institute for Innovation in Education, Hebrew University						
Wechsler IQ test for preschoolers—WPPSI III	2:6–7:3	A	Psychological Counseling Service	USA + UK				N=1700	Intelligence	
Wechsler IQ test for preschoolers—WPPSI	4:0–6:5	A	Psychological Counseling Service	USA		Yes	Informal translation	N=1072	Intelligence	

Test	Age	A: 3–6 B: 6–9 A+: 3–9	Education Ministry dept. that uses it	Developer	Year published	Hebrew translation and norm sample (N)	Arabic translation and norm sample (N)	Foreign norms (N)	Aspect of functioning tested	Emotional/behavioral
Guralnik screening test of language skills of Hebrew-speaking preschoolers	2:7–6:0	A				Yes			Language skills	
Miller Assessment for Preschoolers (MAP)	2:9–5:8	A		USA				N=1200	Sensorimotor abilities, cognitive skills, and complex tasks requiring a combination of the two	
House-Tree-Person (HTP)	Preschool	A	Psychological Counseling Service						Projective test of emotional functioning	*
Children's Apperception Test (CAT)	Preschool	A	Psychological Counseling Service						Projective test of emotional functioning	*
Phonological awareness test as a predictor of reading acquisition (Tubul, Lapidot, & Wohl, 1995)	Kindergarten	A				N=97			Language skills	
Bayley Scales of Infant Development, 2nd ed.	1 month–3.5	A		USA				N=1700	Cognitive/mental	*
Bayley Scales of Infant and Toddler Development, 3rd ed.	1 month–3.5	A		USA				N=1700	Cognition, language, motor skills, adaptive behavior, and socio-emotional abilities	*
Esti Guralnik's screening test, standardized for preschoolers	Preschool	A	Special Education						Evaluation of communication, language, and speech	
Irit Katzenberger's diagnostic test, inappropriate for complex populations	Preschool	A	Special Education						Evaluation of communication, language, and speech	

Test	Age	A: 3-6 B: 6-9 A+: 3-9	Education Ministry dept. that uses it	Developer	Year published	Hebrew translation and norm sample (N)	Arabic translation and norm sample (N)	Foreign norms (N)	Aspect of functioning tested	Emotional/ behavioral
Stanford-Binet Intelligence Scales, 5th ed.	2-85	A+B		USA		N=380		N=4800	Intelligence	
Kaufman test for children (K-ABC)	3-12	A+B	Psychological Counseling Service			N=2000	Yes		Intelligence	
Ma'ase test of spoken-language processing (Romm & Morag, 1999)	5-12	A+B				N=686		Yes	Vocabulary, ability to organize words in large groups	
Developmental Test of Visual-Motor Integration (VMI, Beery)	2-18	A+B		USA		N=11000		Yes	Visual perception, motor and visual- motor coordination	
Bender Visual- Motor Gestalt Test	3-adult	A+B						Yes	Identifying visual- motor difficulties	
Test of Visual- Perceptual Skills (n- m), Revised (Gardner)	4:0-12:11	A+B						N=1032	Visual perception (non-motor): visual discrimination, visual memory, visual spatial relationships, form constancy, visual sequential memory, figure- ground, visual closure	
Motor-Free Visual Perception Test, 3rd ed. (MVPT-3)	4:0-11:11	A+B						Yes	Visual perception (non-motor): spatial relationships, visual discrimination, figure-ground, visual closure, visual memory	

Test	Age	A: 3–6 B: 6–9 A+: 3–9	Education Ministry dept. that uses it	Developer	Year published	Hebrew translation and norm sample (N)	Arabic translation and norm sample (N)	Foreign norms (N)	Aspect of functioning tested	Emotional/behavioral
Maslul: screening test for learning disabilities—adaptation of the American DST	7:6–16:5	A+B		PsychTech Israel and Bar-Ilan University		Hundreds of pupils			Motor skills and basic language skills	
Shatil test for early identification of difficulties in acquisition of reading and spelling: mapping test for preschool and first-grade teachers	Preschool and first grade	A+B		Israel				Unrepresentative sample	Early identification of learning disabilities	
Vineland Adaptive Behavior Scale	1 month–18 years	A+B					Yes, informal		Test of adaptive functioning: communication, day-to-day functions, socialization, motor skills (up to age 5)	
RAN	Preschool and first years of school	A+B		USA					Lexical retrieval	
Shefi Cohen Mashiah	Preschool and first years of school	A+B	Special Education						Naming	
Preschool Language Scale-3 (PLS-3UK)	2 weeks–6:11	A+B		USA			Informal translation		Listening comprehension, expressive communication	

Test	Age	A: 3–6 B: 6–9 A+: 3–9	Education Ministry dept. that uses it	Developer	Year published	Hebrew translation and norm sample (N)	Arabic translation and norm sample (N)	Foreign norms (N)	Aspect of functioning tested	Emotional/behavioral
Stanford-Binet Intelligence Scale, 4th ed.	2–adult	A+B		USA				Yes	Verbal reasoning, abstract/visual reasoning, quantitative reasoning, and short-term memory. Psychologists also use it as a projective test even though there are no norms and it is not valid as a projective test.	
Wechsler test Wisc-R-95	6–12	B	Psychological Counseling Service			Yes	Arabic translation and standardization in progress		Intelligence	
“What Next”: Education Ministry Publications Dept., Pedagogical Administration, Dept. for Elementary School Education	Grades 2–6	B	Elementary School Education						Reading levels: story comprehension, reconstruction of a story in writing, basic reading skills	
Marie Clay’s dynamic diagnosis using a narrative text	Grades 2–6	B	Elementary School Education	Dept. for Elementary School Education					Attitudes toward reading, reading orientation, reading fluency, reading accuracy, story comprehension, etc.	

Test	Age	A: 3–6 B: 6–9 A+: 3–9	Education Ministry dept. that uses it	Developer	Year published	Hebrew translation and norm sample (N)	Arabic translation and norm sample (N)	Foreign norms (N)	Aspect of functioning tested	Emotional/ behavioral
Tests of writing in various genres	Grades 2–6	B	Elementary School Education	Dept. for Elementary School Education, Beit Berl Language Skills Center—Dr. Zvia Walden, Anna Sandbank, and others					Writing level in terms of content, structure, and language, ability to write a story, and other common sub-genres	
Ma'akav kit	Grades 2–6	B		Mofet Institute, Nitzan, Dr. Michal Shani		Representative countrywide sample, N=400			Reading and writing	
From Aleph to Tav	Grades 2–6	B		Mofet Institute, Nitzan, Dr. Michal Shani		Representative countrywide sample, N=400			Mechanisms underlying reading disabilities	
Dynamic diagnosis using an informative text	Grades 3–6	B	Elementary School Education	Dept. for Elementary School Education, Educational Involvement Division					Using a strategy for reading a nonfiction text, extracting information from an informative text	
Elul test kit	Grades 1–12	B		Univ. of Haifa research team headed by Prof. Zvia Breznitz		Representative countrywide sample, N=1500	A translation exists.		Basic skills: reading one's native language, mathematical thinking, foreign-language acquisition, thinking skills	
Reading and writing test for first grade	First grade	B	RAMA	National Institute for Testing and Evaluation and Dr. Michal Shani		N=139	Translation in progress		Reading, writing, and reading comprehension skills	

Test	Age	A: 3-6 B: 6-9 A+: 3-9	Education Ministry dept. that uses it	Developer	Year published	Hebrew translation and norm sample (N)	Arabic translation and norm sample (N)	Foreign norms (N)	Aspect of functioning tested	Emotional/behavioral
Achievement test on acquisition of reading and writing in second grade (as part of the GEMS tests)	Second grade	B	RAMA	National Institute for Testing and Evaluation and Dr. Michal Shani		N=16,000	A translation exists.		Reading, writing, and reading comprehension skills	
Mabatim		A	Pre-School Education						Language, motor, cognitive, and behavioral	*
Ma'agan										
Hatzav: inclusive evaluation of pupils with moderate to mild mental retardation										
Shevilim: inclusive evaluation of pupils with moderate to mild mental retardation										
Legani: evaluation in special education preschools and inclusion preschools in Israel				Rachel Shahr, under the supervision of Nira Kadman					All aspects of functioning	
Social Circles			Special Education	Isaac Israel					Communication, language, and speech	
ITPA			Special Education				Yes		Verbal analogies only	*
Achenbach		A+B	Psychological Counseling Service	Achenbach		Translated. There are norms for young children.	No	Yes, in the US	Evaluation of psychopathology, divided into internalized and externalized problems	*
Connors		A+B	Psychological Counseling Service	Connors		Translated	No	Yes, in the US	ADHD evaluation	

Continuation of table:

Test	Math	Language	Motor/ sensory	Intelligence	Other cognitive	Personnel authorized to administer it	Setting	Purpose	Referrer	Recipient of findings and method of follow-up, if any
HATAF										
PACT										
HA'ETGAR										
Wechsler IQ test for preschoolers—WPPSI III				*			Individual			
Wechsler IQ test for preschoolers—WPPSI				*		Psychologists only	Individual			
Guralnik screening test of language skills of Hebrew-speaking preschoolers		*				Communication therapists and language development experts	Individual			
Miller Assessment for Preschoolers (MAP)			*		*	Occupational therapists	Individual			
House-Tree-Person (HTP)							Individual			
Children's Apperception Test (CAT)							Individual			
Phonological awareness test as a predictor of reading acquisition (Tubul, Lapidot, & Wohl, 1995)		*					Individual			
Bayley Scales of Infant Development, 2nd ed.					*	Specially trained psychologists only	Individual			
Bayley Scales of Infant and Toddler Development, 3rd ed.		*	*		*	Specially trained psychologists only	Individual			
Esti Guralnik's screening test, standardized for preschoolers		*								
Irit Katzenberger's diagnostic test, inappropriate for complex populations		*								
Stanford-Binet Intelligence Scales, 5th ed.	*				*	Psychologists	Individual			
Kaufman test for children (K-ABC)				*		Psychologists only	Individual			
Ma'ase test of spoken-		*				Communication	Individual			

Test	Math	Language	Motor/ sensory	Intelligence	Other cognitive	Personnel authorized to administer it	Setting	Purpose	Referrer	Recipient of findings and method of follow-up, if any
language processing (Romm & Morag, 1999)						therapists, special education teachers, psychologists				
Developmental Test of Visual-Motor Integration (VMI, Beery)					*	Occupational therapists	Individual/ group			
Bender Visual-Motor Gestalt Test			*		*	Psychologists and occupational therapists	Individual/ group			
Test of Visual-Perceptual Skills (n-m), Revised (Gardner)					*	Mainly occupational therapists, but also psychologists in the Educational Psychology Service				
Motor-Free Visual Perception Test, 3rd ed. (MVPT-3)					*	Occupational therapists in pediatric units	Individual			
Maslul: screening test for learning disabilities— adaptation of the American DST		*	*			Professionals in the education system: guidance counselors, special education teachers, communication therapists, didactic diagnosticians, specially trained teachers, and educational and clinical psychologists				
Shatil test for early identification of difficulties in acquisition of reading and spelling: mapping test for preschool and first-grade teachers						Preschool teachers, schoolteachers, and specialists in psychology or learning disabilities	Individual			
Vineland Adaptive Behavior Scale			*	*	*		Individual			
RAN		*								
Shefi Cohen Mashiah		*								

Test	Math	Language	Motor/ sensory	Intelligence	Other cognitive	Personnel authorized to administer it	Setting	Purpose	Referrer	Recipient of findings and method of follow-up, if any
Preschool Language Scale-3 (PLS-3UK)						Communication therapists				
Stanford-Binet Intelligence Scale, 4th ed.	*	*			*	Psychologists	Individual			
Wechsler test Wisc-R-95				*			Individual			
“What Next”: Education Ministry Publications Dept., Pedagogical Administration, Dept. for Elementary School Education		*		—			Group	Informal reading evaluation, accompanied by instructions for designing a didactic intervention program, self-test tool, class survey—tool for preliminary sorting of the class into reading levels		
Marie Clay’s dynamic diagnosis using a narrative text		*					Individual			
Tests of writing in various genres		*					Individual and group			
Ma’akav kit		*				Teachers or psychologists	Individual			
From Aleph to Tav		*				Professional diagnosticians of learning disabilities	Individual			
Dynamic diagnosis using an informative text		*					Individual			
Elul test kit	*	*				Teacher supported by a learning evaluator	Group			
Reading and writing test for first grade		*				Teachers	Individual			
Achievement test on acquisition of reading and		*					Group			

Test	Math	Language	Motor/ sensory	Intelligence	Other cognitive	Personnel authorized to administer it	Setting	Purpose	Referrer	Recipient of findings and method of follow-up, if any
writing in second grade (as part of the GEMS tests)										
Mabatim		*	*		*	Preschool teachers trained by teams from the Psychological Counseling Service and by instructors from the Dept. for Pre-School Education	Individual	Preliminary identification of difficulties for purpose of interventions and the design of tailored intervention plans		
Ma'agan										
Hatzav: inclusive evaluation of pupils with moderate to mild mental retardation										
Shevilim: inclusive evaluation of pupils with moderate to mild mental retardation										
Legani: evaluation in special education preschools and inclusion preschools in Israel						Preschool teachers				
Social Circles		*								
ITPA		*								
Achenbach										
Connors										