The International Education Journal: Comparative Perspectives Vol. 17, No. 4, 2018, pp. 129-146.
https://openjournals.library.sydney.edu.au/index.php/IEJ

What more can we learn from PISA tests? A comparative analysis of the long-term dynamics of Israeli international educational achievements

Yosi Yaffe
Ohalo Academic College, Katzrin, Israel.
Tel-Hai Academic College, Qiryat Shemona, Israel: *yaffeyos@telhai.ac.il*
David Burg
Shamir Research Institute, University of Haifa.
Ohalo Academic College, Katzrin, Israel: *davidburg@ohalo.ac.il*

One of the most prominent comparative international tests in recent years is conducted by the Programme for International Student Assessment (PISA), which is used to assess students' literacy in Reading Comprehension, Mathematics, and Science. Israel has taken part in these assessments from its inception, and since then it has been gradually dropping in the world ranking. By 2015, Israeli students were located between 37–40 out of 70 countries, down from 31–33 in 2000. Using Israel's data in the PISA cycles, we aim to offer an alternative comparative view of these international results, based on measures of self-change. These measures show a consistent and significant system-wide improvement of the Israeli educational performance in all three PISA tests since 2000. Not only does this contrast with its downward trend in the international ranking, which stems from the inclusion of more countries, but the rate of change was consistently positive over time and surpassed that of its European and American counterparts. Indeed, these measures of change show a general growth of about 23 points over 15 years, which reflects a consistent improvement of 6 to 9 points in average per cycle since the 2000 results. In light of these findings, the paper discusses the significance of measurements of change in comparative international tests.

Keywords: PISA, International tests, Education, Achievements, Measurement of change

INTRODUCTION

Programme for International Student Assessment (PISA)

It is common to distinguish between two main types of student assessment: large-scale assessment and classroom assessment. Large-scale assessments test a large number of students and are a means of obtaining information about the relationship between various educational achievements and the factors that may explain them (the Committee for Measurement and Evaluation in Education, 2005), such as curriculums, pedagogies, as

well as the quality and functioning of an educational system. In Israel, large-scale assessments of achievements are carried out through two main systems: Matriculation (*Bagrut*) and a system of measures of efficiency and school growth (*Meitzav*). The latter constitutes a system of social-educational indices designed to provide an up-to-date picture of the educational system at its various levels, based on national assessments in Native Language, Mathematics, English, and Science. At the international level, large-scale assessments are carried out using comparative achievement tests such as that implemented by the Programme for International Student Assessment (PISA).

Such comparative international tests have been introduced in the world since the 1960s. Since the 1990s, the use of these tests has expanded greatly with the support of international organizations such as UNESCO, the World Bank, and the Organization for Economic Co-operation and Development (OECD). The results of these tests attract considerable public attention among countries, with most of the them dissatisfied with their international position even when located at the top of the world ranking (Atkin & Black, 1997; Blass, 2016). One of the most prominent comparative international tests since the early 2000s is the PISA, which is used to assess academic literacy in reading comprehension, Mathematics, and Science among pupils in the 15-year-old age group. The objective of PISA assessments is to test students' ability to use their literacy skills to understand and interpret a variety of written materials they will likely meet during their lives, their ability to apply their mathematical knowledge to tackle various real-world math-based problems, and their ability to use their scientific knowledge to comprehend and interpret diverse scientific situations and challenges (Turner & Adams, 2006). According to the OECD (2016), the PISA tests are designed to serve as an indicator of "students' capacity to apply knowledge and skills in key subjects, and to analyse, reason and communicate effectively as they identify, interpret and solve problems in a variety of situations" (p. 15) (this refers to the concept of literacy). PISA also indicates the relevance of these students' capabilities for lifelong learning, "as (it) asks students to report on their motivation to learn, their beliefs about themselves and their learning strategies."

PISA tests have differed in terms of their disciplinary focus in each cycle, with the emphasis on reading literacy in 2000, Mathematics in 2003, Science literacy in 2006, Reading Comprehension in 2000, and Mathematics again in 2012 (Blass, 2016). The sample includes at least 150 schools in each country, with a minimum of 5,250 students randomly sampled. In addition to the disciplinary literacy tested by PISA tests, students are asked to provide demographic information, educational and professional preferences, personal characteristics, and characteristics of their family environment (Turner & Adams, 2006).

Performance on PISA tests are standardized, with an overall mean score of 500 points, and a standard deviation of 100 points. Between 2000 and 2015, the leading OECD countries with the highest scores in these tests were alternately: Finland, Canada, Chinese Taipei, Japan, China (Shanghai), Singapore, Australia, and New Zealand (not necessarily in this order), and sometimes also European countries such as Estonia, the Netherlands, and Liechtenstein. PISA average scores of the three disciplines in these levels over the years approximately ranged from 530 to 555 points, between one-third and one-half of the standard deviation above the average. Some countries have been observed to reach even higher, such as China's (Shanghai) 2012 average score exceeding 580. There generally appears to be a correlation between achievements in the three disciplines tested

among these leading countries, that is they tend to achieve relatively high scores in all three subjects.

Dozens of countries have begun to evaluate their students' abilities based on their achievements in comparative international tests, based on the assumption that a high ranking in these tests reflects a better future national ability to compete in the global market (Rochex, 2006). As a result, the popularity of these exams has begun rising in recent decades. The results are widely covered in the media, with attempts to analyse and explain the relative achievements in each country (Feniger, Livneh, & Yogev, 2012). Most explanations offered by the experts around the world regarding students' performance were directly attributed to the quality of the countries' educational system in terms of structure and policy (Feniger & Lefstein, 2014). After its students were ranked below the OECD average in 2000, Germany enacted policy to implement comprehensive reforms in the education system, including the establishment of an educational standards monitoring mechanism. The United States (US0 uses state-level standardized testing, and also has national testing regimes. Following these examples, Israel introduced a policy of setting standards, including evaluation measures for scholastic achievement, school efficiency, and growth (*Meitzav*) (Yogev, Livneh, & Pniger, 2009).

However, studies published in recent years suggest another way of examining and understanding the international variance in student achievement on the PISA tests. Feniger and Shavit (2011), for example, showed that the relatively high birth rate in Israel can explain most of the discrepancies between the Israeli students' achievements and the PISA's international mean scores in 2000 and 2006. This high birth rate is reflected in the formation of larger families and crowded classes, which, in turn, affect children's educational achievements. Moreover, studies show that high birth rates are associated with a decline in cognitive development and educational achievement first and foremost through family size, beyond the relative reduction in national student investment *per se*. This is due to the tendency for children to receive a smaller share of family resources in families with more siblings, including less parental attention and involvement in the child's education (Blake, 1989). Further, an additional analysis of the PISA scores from 2000, conducted in Israel in 2002, showed that students' achievements in Israel are almost entirely predictable by two macroeconomic variables: Gross Domestic Product per capita (GDP) and demographics, notably the proportion of young population (R^2~0.80, Feniger et al., 2012). The researchers pointed out that the economic and demographic elements of the countries largely dictate their success in the PISA tests, rather than the investment in education and the quality of the educational system. Yogev et al. (2009) concluded in this context that:

> The low ranking of the Israeli students among the other countries does not reflect such poor achievements in the education system that justify a public panic response. It is not at all clear whether these findings should lead to the conclusion that there is a decline in academic achievement in Israel. There may be such a decline, and there may not. In any case, our findings indicate that the results of the international tests do not provide sufficient grounds for the claim regarding the Israeli decline in achievements. (p. 349)

Economic and demographic factors can similarly predict PISA scores among other countries (Feniger et al., 2012), questioning its validity as an indicator of the educational abilities of students around the world and the quality of their national education systems. For example, the PISA reports from 2000 revealed a close connection between mothers'

education and students' achievements in all three PISA's disciplines, which reflects more than half a standard deviation gap compared with those students whose mothers did not complete high school (Anderson, Lin, Treagust, Ross, & Yore, 2007).

Assessing learners' achievements by measurement of change

A standard way to evaluate performance is based on equivalent comparison of the performance level between individuals or groups. This approach is common and widely accepted for the purpose of comparing achievements among individuals and groups in the fields of sports, education, and more. In the field of education, this approach is used by decision-makers to classify and scale learners according to their achievements on a seemingly equivalent performance scale. On a national or international scale, these assessments are used to compare the achievements of many students in order to identify patterns of strength and weakness in teaching methods, pedagogies, and educational systems, as well as to assess accomplishment of educational goals (National Research Council, 2003). Nevertheless, assessing achievement through an inter-learner comparison at a given time point is not sufficiently sensitive to the dynamics of individual learners' change relative to their previous state (e.g., when assessing the effectiveness of new curriculums) and hence does not necessarily reflect trends of improvement. In some respects, measuring individual change over time is the most important issue in educational measurement, since the main purpose of teaching is to induce learning, and the essence of this change can only be measured by comparing the performance of individuals and groups with their own abilities before and after the learning (Coleman, 1975). The dynamics of the school's achievements, or the educational system at the macro level, argues Coleman, should be the main measure of its educational quality.

A traditional measurement of change deals with two main approaches. One focuses on individual changes over time and addresses the question of whether there has been an improvement in performance, whether the improvement is stable or gradual, and whether its trend straightens at the end of a certain period. The second approach deals with the question of groups' differences in nature and rate of change, so that one group improves faster than the other. The first method usually precedes the second method, since it is initially necessary to identify the individual's pattern of change prior to evaluating whether one pattern of change is systematically different from that of another individual (Willet, 1994). Barrett and Alexander (1989) offer three operational definitions of measures or criteria for change: 1) long-term changes in the average level of individual or group performance; 2) changes in the correlation of validity metrics over time; and 3) changes over time in rankings of individuals within a group. It has been suggested in recent decades that educational measurement of individual change should shift from a two-wave measurement (e.g., pre-post testing) to the measurement of continuous and multiple growth cycles (Bock, 1976; Willett, 1994). Measuring the continuous multiple change in the average level of performance will be adopted in this study for testing Israeli self-change in PISA tests overtime.

Evaluation of international achievements in PISA tests in terms of self-change

A brief review of the PISA results of the various countries over the years since its onset indicates the dominant way of evaluating the students' level and the educational quality in each participant country. Thus, an international comparison of the nations' average score in each discipline is used by public opinion and educational policy makers to rate

their national educational system relative to the rest of the participating countries. Along with this traditional comparative view of international achievements in education, there is also an equally important perspective for evaluating achievements on PISA tests. This refers to the trend of change (improvement or deterioration) that has occurred in the countries' achievements during the past decade-and-a-half of PISA tests, beyond their achievements *per se* in each cycle. This perspective has attracted much less interest and attention among experts and stakeholders. Even prior to the PISA era, Willett (1989) warned about the tendency of experts in educational measurement to assess the efficiency of different pedagogies by comparing learners' educational performance at a single point in time rather than in terms of self-growth over time. Within-subject change in PISA scores over time—that is in this context within each country—may not only provide forecasts of countries' achievements in these tests for the upcoming years but may also constitute an important measure of the effectiveness of the educational processes at both the micro- and the macro-levels (school and national education, respectively). The scant consideration given to this perspective in assessing international achievements shows a surprising trend regarding Israel, whose students were among the top ten countries in the world in terms of improvement in international comparative tests over time, including in the PISA (Blass, 2016).

In light of the substantial difficulties involved in educational comparison of students from different nationalities and cultures, as pointed out by various experts in the past (Bonnet, 2002; Goldstien, 2004; Romainville, 2002), evaluating the PISA results via the measures of self-change in achievements over time (i.e., improvement or decline against country's own past scores) seems fundamental. International comparative tests that purport to be educationally equivalent (in a sense of measuring differences in performance related to identical educational skills and contexts) must give more consideration to dynamics of self-change in achievement over time. This is because indicators of change within nation as measured by PISA results are a type of standardized measure whose values allow for a more equal comparison between the performance of students of different cultures and languages.

The differences between the Israeli scores and the other countries' scores in Mathematics and Science over the years has led to the conclusion that the educational state of the Israeli student has worsened due to the deteriorating condition of the Israeli educational system in recent decades (e.g., Ben-David, 2003; 2010). However, the decline of the Israeli students in the international ranking in PISA tests does not necessarily reflect diminished educational capabilities compared to their previous performance. In line with Yogev and colleagues (2009), we aim to introduce here a method of examining the performance of an educational system in terms of self-change as a function of time in PISA tests. While PISA provides a trend analysis in the primary analyses they publish after each round of (i.e., it reports the change relative to the last cycle), this paper endeavours to generate a timeseries analysis of long-term PISA data, resulting in a quantified comparable score of the countries' changes over time. Considering the relatively low attention given to this perspective on countries' achievements in international tests, the current study strives to provide a complementary analysis of the change in PISA results relative to these countries over all the years.

By studying the case of Israel, the perspective presented here may yield new insights regarding the educational state of various countries and national groups from an international perspective.

The study's objectives

The study intends to show that from 2000 to 2015, not only did the achievements of Israeli students in the three evaluated educational disciplines not decrease, they, indeed, improved consistently and significantly, even though their international ranking among the OECD countries declined. In this paper, we will attempt to implement the orientations of a self-change measurement discussed above in order to assess Israel's individual improvement in the PISA tests over the years and compare them with that of several selected countries. Examination of Israeli students' achievements by using measures of change will show a consistent trend of improvement in all three disciplines. Considering the relatively low attention given to this perspective on countries' achievements in international tests, the current study strives to provide a complete analysis of the change in PISA results relative to these countries over all the years.

In order to analyse Israel's case in terms of the measurement of change, the following countries were selected for comparison purposes: Finland (FL), the United States (US), Britain (UK), and France (FR). Being a world leader in the PISA international ranking, Finland has been viewed by many in Israel (IL) and globally for a decade-and-a-half as a focus of interest and a role model educational system. UK, FR, and the US (in order) are currently located at centre or above of the PISA international ranking, with at least two cycles ranked in the top 15. These countries were chosen due to similarities in ethnic ratios between demographic majorities and major minorities. Also, their economics (GDP) are similar (though the US is higher), with comparable unemployment (except for FR). Interestingly, UK and FR are much less diverse than the US, which is similar to IL. Finally, these countries have similar liberal and advanced democratic forms of government. FL was included here for its position as "education powerhouse," even though its investment in education is relatively higher and it has a homogenic culture.

Despite cultural and educational similarities with IL, these countries are still located significantly higher in the international rankings of PISA tests, despite the Israeli effort to imitate other nations' policies, such as "No Student Left Behind" in the US (Yogev et al., 2009). Nevertheless, these three countries have experienced an ever-downward trend, while IL has been continuously improving. The aim of this study in this context is to express the differences between these countries in terms of self-change measurements and to show how this view of achievements reverses the hierarchical structure between them.

METHODS

Variables

The results of PISA tests in the three literacy disciplines: Reading, Mathematics, and Science, between 2000 and 2015 (see Table 1 for details), as reported for the following five countries: FL, IL, US, UK, and FR.

Data source

This work is based on the data of the achievements recorded for the five selected countries participating in the PISA evaluations, as published in publicly available online material provided by the OECD) 2003, 2005, 2007, 2010, 2014, 2016). With respect to each

country, the national average scores for each PISA test in Science, Mathematics, and Reading Comprehension were isolated and analysed comparatively.

RESULTS

Distribution and ranking of international PISA scores (2000–2015) from a conventional perspective

Here we review the achievements of the five discussed countries in PISA tests over the past two decades, providing their published international rankings over the years. In order to establish a concise comparative perspective, this report will focus on three-time points: 2000, 2009, 2015. As mentioned earlier, PISA scores are standardized to a mean score of 500 and standard deviation of 100.

Table 1 shows the scores and rankings of the five countries in three PISA tests from 2000 to 2015. An overview of the PISA test data in its first cycle (2000) shows that FL ranked at the highest place among the five countries and shared one of the five top places of the total world ranking in all three disciplines. The US is ranked between the 15th to the 20th place in this PISA cycle, with better achievements in Science. Ultimately, the two European countries, UK and FR, achieved higher rankings than their US counterpart and are ranked between 5-8 (UK) and 11–13 (FR), while IL is ranked far behind, between the 31st and 33th places. In 2003, FL preserves its international ranking in the top five, the other three countries lose approximately three to five places from their former international ranking in PISA tests, and IL did not take part in this cycle.

Table 1: Achievements data of the 5 selected countries in PISA tests between 2000–2015

Country	**Test**	**(Ranking) & score**					
Participating countries (number)		2000 *41*	2003 *41*	2006 *56*	2009 *64*	2012 *65*	2015 *70*
1. FL	Reading	(1) 546	(1) 543	(2) 547	(3) 536	(6) 524	(4) 526
	Math	(5) 536	(2) 548	(2) 548	(5) 541	(12) 519	(5) 532
	Science	(4) 538	(1) 548	(1) 563	(2) 554	(5) 545	(5) 531
2. UK	Reading	(8) 523	(12) 507	(18) 495	(25) 494	(23) 499	(21) 498
	Math	(9) 529	(17) 510	(25) 495	(28) 492	(26) 494	(15) 507
	Science	(5) 532	(12) 518	(14) 515	(16) 514	(21) 514	(15) 509
3. FR	Reading	(15) 505	(18) 496	(23) 488	(22) 496	(21) 505	(20) 499
	Math	(11) 517	(13) 519	(24) 496	(22) 497	(25) 495	(27) 492
	Science	(13) 500	(14) 511	(25) 495	(27) 498	(26) 499	(27) 495
4. US	Reading	(16) 504	(19) 495	(18) 495	(17) 500	(24) 498	(24) 497
	Math	(20) 493	(30) 477	(36) 474	(31) 487	(36) 481	(25) 494
	Science	(15) 499	(24) 491	(29) 489	(23) 502	(28) 497	(25) 496
5. IL	Reading	(30) 452	-	(40) 439	(37) 474	(34) 486	(37) 479
	Math	(31) 433	-	(41) 442	(42) 447	(41) 466	(40) 470
	Science	(33) 434	-	(38) 454	(42) 455	(41) 470	(40) 467

The following will focus on the main trends that have emerged since 2009, given the sharp international increase in participation in PISA tests, from 30 to 64 countries. In the 2009 cycle, FL retains the 2nd–5th positions, the US ranges from 17-31 (preference to Reading), while UK and FR rank 16-25 (preference to Reading) and 22–27 (preference to Math and Science), respectively. IL is located far behind, between the 37th and 42th places (preference for Reading). Against the background of the increasing number of participating countries, it can be seen that all of the countries discussed here (apart from FL, as mentioned) lost their starting position in the global rankings. It should be noted that the ranking gaps between the three disciplines consistently expanded for all these countries, concomitant with the expansion of the included nations.

An overview of the last cycle of PISA tests so far, the last of which was conducted in 2015, shows again that only FL maintained its position in the top five places in the league table. The US occupies the 24th-25th places, narrowing the ranking gaps between the disciplines, but apart from Mathematics it does not achieve significant progress. UK (15–21) and FR (20–27) achieve a general improvement of several places each, while

It is unclear what can be learned from the trends reviewed so far regarding the dynamics in the achievements of the five countries discussed here. In general, it appears that since 2003 the five countries have occupied a consistent range of rankings in the international PISA tests (subject to changes in the number of participating countries), with a consistent tendency since 2000 to lose a few places in each cycle, except for FL, and to stabilize or improve in the last cycle. Table 1 generally ranks the five countries according to their mean scores on PISA tests and their place in the overall international rankings over the years. Accordingly, FL is ranked first among the top five, followed by UK, FR, and, finally, the US and IL with the lowest scores, with most of them below the OECD average.

Examining the achievements data among the five countries discussed here essentially shows that the general hierarchy is also evident at the individual level in each of the subjects (Figure 1). With respect to all three disciplines, FL leads the five with above-average scores over the years, and IL is in last place with below-average achievements. Regarding Mathematics, UK and FR appear to be in the second place with above-average achievements in most years, while the US is lower with below-average scores. With respect to science, UK is in second place with above-average achievements in most years, while FR and the US are approaching the international average (with a slight advantage for FR). As for Reading, finally, it can be seen that most of the three middle-table countries achieve quite similar achievements in this field over the years, with scores that approaching the international average.

The conventional international comparison of the achievements of the five countries over the years offers a clear hierarchy for at least three of the five countries in terms of the educational level of their students in the three disciplines. Accordingly, FL's achievements in the PISA tests clearly reflect the highest average performance in Reading, Mathematics, and Science literacy among the five, followed by the European countries, the US, and, finally, IL.

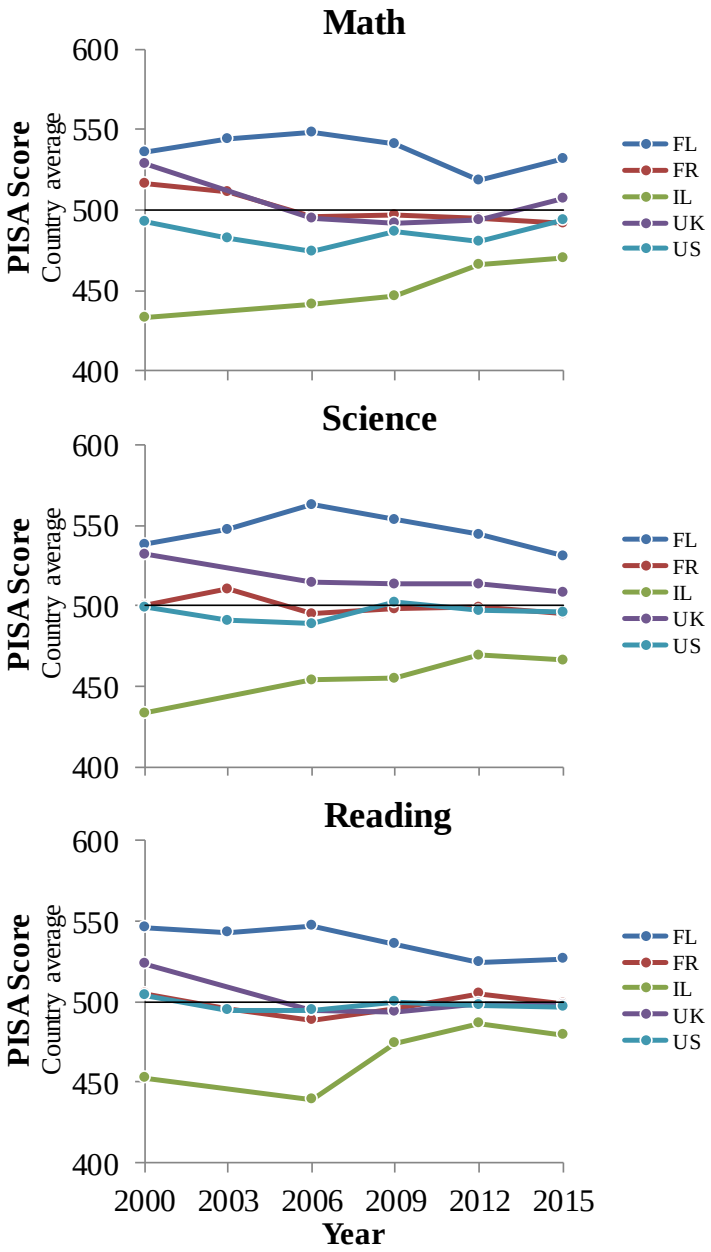

Figure 1: Description of the 5 selected countries in PISA tests between the years 2000–2015

Evaluating the results of PISA tests using measurements of change

Here we will examine the achievements of the five selected countries in terms of change (i.e., increase or decrease) in each PISA cycle compared with previous achievements, in order to rank the countries according to an overall change-over-year index, that describes the trend (positive / negative) and its size.

Regarding the five countries discussed here, the differences in achievement relative to the 2000 results in the three PISA tests are calculated. A positive change of a given country's achievements is described in points relative to the starting point of the PISA test in 2000. Figure 2 shows that IL is the only country among the five that shows a constant positive upward trend since 2000 in each discipline, except for Reading in 2006. Not only is the Israeli trend of change generally positive (i.e., the scores' differences between a given

year and 2000 is larger than zero), but also the increment relative to 2000 for the most part grows from one cycle to the following. Following the recent PISA tests, the Israeli improvement trend grew to approximately 40 points, an improvement of approximately 8% in the three subjects compared with its starting point in 2000. Finland is the only country, except IL, whose achievements over the years in Mathematics and Science are partially in an upward trend relative to 2000, although as of 2006 the rate of its improvement between the following years steadily decreased. This downward trend is consistent, showing FL's achievements in Mathematics and Science in recent PISA tests falling below their levels in 2000.

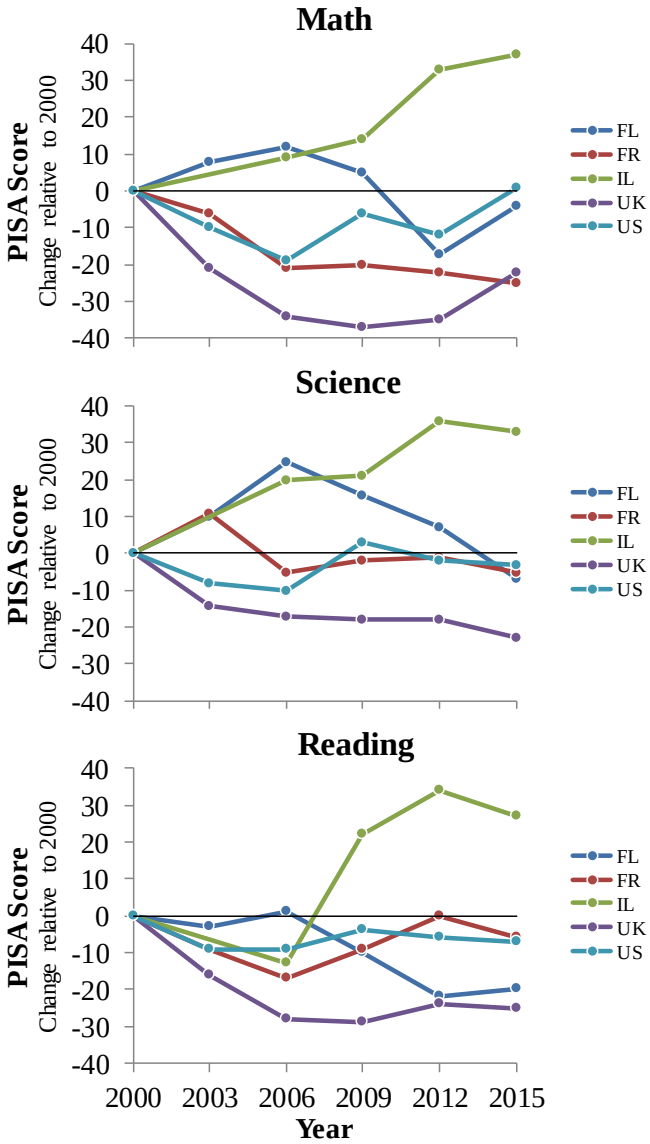

Figure 2: Description of the scores' change in the achievements of the 5 selected countries in the PISA 3 tests between 2000 and 2015

A downward trend is observed for the other countries, in achievements in all PISA tests since 2000 (as reflected in the scores below the midline in Figure 2), with this trend remaining relatively stable and not deteriorating. The US, for example, has maintained a

stable gap of a few points over the years, compared with its 2000 achievements in the three disciplines. UK, which began with fairly high achievements in 2000, experienced a sharp decline in all the disciplines in 2006, but this decline stabilized with no further deterioration. A similar trend emerges for FR regarding Reading and Mathematics, but since 2006 its achievements in sciences are not substantially lower than in 2000.

In order to weigh the dynamics of achievements of the five discussed countries into a general measure of change, we calculated the mean changes in each PISA discipline compared to the 2000 achievements as a reference point (i.e., the mean difference between the results in 2000 and the results in all subsequent cycles). Generally, the higher this index is, the more positive trend of change it reflects in each discipline compared to PISA's starting point in 2000. The change in the three disciplines was also calculated. An overall average between-discipline was computed:

$$\bar{D} = \sum_{i=2}^{6} \frac{(X_i - X_{2000})}{N - 1}$$

where *i* is the index of PISA cycles from its second cycle to its sixth cycle and X is the standardized PISA score in the year 2000. This ranks the five countries by the size of their overall change over years

Table 2: Description of the measures of change in terms of the mean differences ($\bar{D}$) between PISA scores in 2000 and the following cycles

Country	Math	Sciences	Reading	Overall change ($\bar{\bar{D}}$)
1. IL	+23.25	+27.5	+17.5	22.83
2. FL	+0.8	+10.2	-10.8	0.06
3. US	-7.0	-10.4	-4.0	-7.13
4. FR	-18.8	-0.4	-8.2	-9.13
5. UK	-29.4	-18.0	-24.4	-23.9

Table 2 shows that the only two countries with an overall positive change are IL and FL. These measures reflect a general trend of improvement relative to these countries' scores in the PISA tests over the years. With respect to FL, the data indicate an average decrease of 11 points in the Reading scores, compared with its starting point achievement in PISA 2000. In contrast, IL accomplished a larger improvement in comparison with FL in all three subjects, with an average gain of nearly 23 points in its achievements over the years (compared with the starting point in 2000). This trend is reflected in Figure 2, (see L), which is mostly above the midline referring to the three disciplines. In contrast, the overall measure of change is negative for the other countries, reflecting an average decline in all three disciplines compared to their starting point in 2000. The stronger negative trend was recorded for UK, which opened the PISA tests with relatively high achievements in all three disciplines (see Table 1) and subsequently deteriorated from this point.

The measure of change in Table 2 properly reflects the dynamics in achievements relative to 2000 as a starting point, but it is not sensitive to the improvement that may occur between two consecutive cycles afterwards. Thus, for example, the improvement trends between two consecutive cycles of UK after 2000 are not reflected in this index, since their achievements in these years are lower than those achieved in 2000. To overcome this problem, the consecutive change measure (CCM) was calculated:

$$\bar{D} = \sum_{i=1}^{5} \frac{(X_{i+1} - X_i)}{N - 1}$$

where *i* is Index of PISA cycles from its first cycle to its fifth cycle; X is the standardized PISA score in cycle *i*. This averages the gaps between the various PISA cycles over the years in all its three disciplines. Table 3 displays these results and ranks them according to the general consecutive change measure, which constitutes a general mean change beyond a discipline. In order to test the significance of the linear change over years (as described by the trends in Figure 1), we calculated the slopes values of the five cycles' scores for each discipline and considered them in relation to the CCM.

Table 3: Description of the measures of change in terms of the mean differences ($\bar{D}$) between PISA scores in cycle *i* and the consecutive cycles

Country	Math	Sciences	Reading	Overall change ($\bar{\bar{D}}$)
1. Israel	+9.25	+8.25	+6.75	+8.08
2. Finland	-0.8	+2.6	-4.0	-0.73
3. USA	+0.2	-2.6	-1.4	-1.26
4. France	-5.0	-1.0	-1.2	-2.4
5. Britain	-4.4	-4.6	-5.0	-4.6

According to the CCM (Table 3), FL manifests a slight negative trend of change whose slope is significant for Reading (β= -0.50, p<.05), while, in general, the countries' ranking does not change in order. The overall CCM index, which reflects the mean change in scores of the following PISA cycles for the three disciplines, places IL at the top of the table, with a higher positive change per year. The Israeli students' achievements improved by more than eight points on average between cycles. The individual measures of change of the Israeli students in Mathematics and scientific literacy were both positive and supported by significant slope of improvement (β=0.34, p<.05 and β=0.39, p<.05, respectively), while in Reading Comprehension they also reached a consistent improvement of approximately seven points per PISA cycle. The positive Israeli changes discussed here are also reflected in the upward trend of the graphs appearing in Figure 2. On the other hand, the other three countries all show negative trends of change between subsequent PISA cycles, with a consistent decline ranging from 1.3 to 4.5 points on average. In this regard, UK showed the strongest negative trend among them, with its students' achievements in all three disciplines dropping by an average of more than four points, reflecting a significant score decline in Science (β= -0.62, p<.05) and Reading (β= -0.37, p<.10). FR's CCM decrease in Mathematics is also supported by a significant negative slope (β= -0.41, p<.05).

In conclusion, two different measures of self-change of PISA scores in its three disciplines place IL at the highest level among the five countries selected for our discussion, even though its ranking in terms of achievements was the lowest since the first cycle. The fact that IL has more room to improve from a lower place than its comparative countries should be taken into account, but probably would not be sufficient for interpreting its advantage on the two measures of change. This is because the achievements of our three middle-table countries since 2003 are themselves approaching

the general average (500 with the standard deviation of 100), and occasionally even lower in all three PISA disciplines (see Table 1).

DISCUSSION

In this article, we suggest a method of comparing the international achievements in PISA tests, based on descriptive measures of change. This approach is based on the orientation of measuring the within-subject change of an individual nation's performance in terms of improvement and growth over time (Barrett & Alexander, 1989; Willet, 1994). Being ranked significantly lower than its counterparts in the international rankings, and generally lower than the global average, IL is a classic case study to illustrate the paper's thesis on the importance of measuring the change in international achievement in PISA tests. IL's performance began in 2000 with achievements below the standard. Though it improved since then (especially in 2009), IL's mean performance is still below the average and remains essentially at that level in 2012 and 2015. However, the main finding that emerges from this analysis using the measures of change proposed here (change over time), is that IL has achieved a positive and consistent improvement in all subjects from 2000 to the present, which is preferable to the trend of change recorded for its European and US peers included in our analysis.

These results, alongside the indexes deriving from the regular comparison between the countries' mean scores on PISA's tests, illustrate the importance of both ways of evaluating countries' educational performance through international tests. This is because one of these indicators alone may, in some circumstance, portray an incomplete and even misleading view of the countries' educational performance over time. The measures of change relative to the achievements of the Israeli students showed an average increase of 6–9 points per PISA cycle over the years, which reflects a statistically significant improvement in Mathematics and Science. As far as IL is concerned, this represents a total improvement of about 23 points in PISA mean scores for the three disciplines of literacy, compared with its achievements in 2000 (i.e., an average improvement of more than 20 points since the starting point, which remains stable over time). Against the background of these indices, the data on the change in achievements of all the other countries was negative, indicating a consistent deterioration in achievements between PISA cycles (which, for three counties was partially supported by a statistically significant trend), and generally compared to the first PISA cycle in 2000. In terms of these proposed measures of change, IL and FL are ranked first among the five countries, while UK is in last place. It should be noted, however, that IL's priority in terms of overall change scores, can be accounted for by the fact that countries who achieve at a high level of performance (i.e., FL) are limited in the extent to which their students are able to improve (a phenomenon known as a ceiling effect), while students from low (or lower) achieving countries have much more room for improvement.

Diverging from the prevailing view of the deteriorating state of IL's educational system (Ben-David, 2003; 2010; Blass, 2016), a conclusion that may be drawn from these findings is that the educational level of Israeli students, as evaluated by the PISA international achievements tests, improved from its first round in 2000 in at least two disciplines, with this improvement remaining stable over the last two decades (i.e., the mean scores of the Israeli students in all the following rounds since 2000 are higher than the starting point). Although it is unclear how significant this improvement is, it is entirely

clear from the findings that there has been no decline in the Israeli students' level as reported by the mean PISA score. Although, IL's performance remains well below the average of participating countries, these results (comparing the international PISA exams throughout all six cycles) reinforce the doubt raised by some researchers regarding the possibility that the Israeli education is constantly declining (Yogev et al., 2009; Feniger et al., 2012). It seems that not only have Israeli students not lost their literacy knowledge in the fields of Science, Math, and Language, but it has also improved considerably over time.

The current work focused deliberately on characterizing this trend of change in descriptive quantitative terms, while analysis of its educational significance is the subject of a separate examination due to its complexity. To the best of our knowledge, a 10-points increase (IL's average increase in Mathematics between PISA cycles) should be considered a significant change within the three years that separate each PISA exam cycle, so that in a cumulative view, this trend amounts to an improvement of 40 points since 2000, which is slightly less than half a standard deviation. Indeed, IL has already been recognized as one of the leading countries in the world in terms of improvement in the international tests (Blass, 2016), although, to our knowledge, this has not yet been analysed throughout the whole cycles. Assuming this, IL's improvement is indeed significant. We suggest, here, that the system of measures of efficiency and school growth (*Meitzav*) introduced by the Israeli educational system in 2002 (right after IL's first participation in the PISA international educational tests in 2000), had something to do with the Israeli students' consistent increase in the performance on the international PISA tests. The Israeli *Meitzav* is a national large-scale assessment which, like PISA, also tests students' performance in the three fundamental disciplines (i.e., Native Language, Mathematics, and Science and Technology). It was introduced as part of a growing worldwide education trend of evaluation to support learning and to increase the accountability of schools and the education system as a whole for teaching and learning.

How can this positive trend of change manifested by the Israeli achievements over the years be reconciled with the fact that its position in the international ranking has declined significantly throughout this period? Thus, for example, between 2000 and 2006, IL achieved nine points increase in Mathematics but dropped ten places in the world rankings. In 2009, the Israeli average in this test grew by five points compared to the previous cycle. However, IL continued to fall one place in the international ranking. Similarly, in Science, the Israeli students improved their achievements by an average of 20 points between 2000 and 2006 but dropped in the world ranking by five places, to the 38th place. Finally, in 2015, IL occupied the 37th-40th ranking, following a significant increase in all subjects since 2006. The main reason for this contradiction between the two trends may be the fact that over the years there has been an appreciable and steady increase in the number of participating countries. It should be noted that the sample size has nearly doubled since 2000, with 15 countries added in 2006 and another eight in 2009. This has, in effect, misrepresented the Israeli improvement in all disciplines and not expressed it in the world rankings. With accurate data in this context shown in Table 1, needless to say, the results achieved by the Israeli students in the last PISA cycle in all three disciplines would place them higher in the international ranking relative to the original 41 countries that participated in 2000.

It is necessary to acknowledge the limitations of international comparisons. In this context, Wuttke (2007) points out some of the statistical obstacles, uncertainties, and lack of transparency inherent in the PISA tests. First and foremost, there are very serious problems relating to the translation of items into the various languages of the participating countries. In addition, some countries are not sampled in a representative way, because the sampling method, in some cases, does not allow representation, as exemplified by the case of China. Thus, for example, the lack of participation of weak students or sectors may boost the country's score by about nine points. Finally, there is also the cultural and ethnic diversity among countries expressed in different attitudes of the educational system, parents, and students towards tests in general and the PISA test in particular. Presumably, in this context, students from different countries have a distinct approach to tests that have little impact on their future (Sjoberg, 2007). In Singapore, for example, "be best – teach to the test" is a mantra, and PISA preparation brochures are displayed in supermarkets for a nominal price. While politicians in the West regard the PISA results very seriously, it is very likely that students do not share their ambitions and, therefore, do not invest the effort required to succeed in a long and difficult task without any benefits. Accordingly, the attempt to measure scholastic competence, involving students' motivation to complete the task successfully, may not succeed. We can cautiously suggest that Israel's cultural attitude and set of values towards examinations, in general, and formal education, in particular, does not allow a valid comparison with other countries.

Thus, an international and multicultural comparison of educational achievements requires greater attention to the achievements in terms of the size of self-change over time. For measurements of change that reflect the individual's progress in academic performance, per student or of the national education system, comparison to itself alone can serve as a basis for an equivalent international comparison of achievements. It is plausibly useless to compare Israel to countries with a fundamentally different demographic and cultural character (such as Finland and China), and even to try to imitate them. Interestingly, over the past decade, when much information has been published regarding the international limitations of the PISA tests, there has been no careful reference by international decision-makers regarding its findings. Therefore, interpreting the results of the PISA tests also in terms of a within national-education-system change may yield more solid insights. This approach assumes that the differences between test cycles in each country are marginal, and that the biases are similar, so it can overcome the international variance that makes it difficult to compare the achievements of countries. Yogev and colleagues (2009) wrote in this context that:

> Emphasizing the countries' grades in a hierarchy that attributes importance to the international stratification of countries according to the achievements of their students is detrimental to the international tests and to the public and educational significance given to them ... We warn against the exaggerated public significance given to the international tests, which sometimes leads to focusing on less substantive educational issues. (p. 350)

They have also been able to point out the huge achievement gaps in PISA scores between different sectors of Israeli society. This explains, to a great extent, the fact that despite the consistent improvement, Israel is still below the international average. This issue, too, amplifies the importance of measuring the self-change over time as a basis for assessing international achievements.

Policy makers who seriously examine the trend of improvement that Israeli students have achieved in the international tests over the past two decades, will ask whether in light of the current demographic and educational conditions one can expect more. Israeli researchers noted that a national expectation of a drastic achievement climb in the international tests is not realistic in light of economic, demographic, and cultural direct and indirect variables that predict these achievements (Feniger et al., 2012). Our findings regarding the Israeli self-change in PISA scores over the years support their positions. Although they reflect a consistent upward trend in educational performance measured by the PISA tests, and although this trend is approaching the global average in another decade, this positive pace does not justify the expectations of the Dovrat committee in 2004 to occupy higher PISA rankings in the international tests in a few years. We. Therefore, propose devoting greater public and educational attention in Israel, as well as in other countries around the world, to the analysis of trends focusing on within-national change in the international test achievements. The educational policy should be more focused on these aspects. In this context, refinement of the indicators for assessing the size of change in international educational achievements, while setting clear and achievable improvement goals for the next decade, at the national level and sectorial level, may prevent the misuse of the positive trend in Israel as it approaches the international average score of PISA.

Our findings do not purport to present a rosy picture of the educational situation in Israel, since we cannot ignore the fact that the constant improvement of Israeli students since 2000 still places them below the global average in all PISA tests. However, it should be noted that a apples-to-apples comparison is not possible due to the structure of the Israeli educational system with multiple sub-systems for each minority demographic. The Arab sector drastically underperforms compared to the secular Jewish students, with nearly no attention given to this disparity. Also, the ultra-orthodox community is completely self-contained and does not conform to Western standards of education. Taken together, it is clear that Israel's international ranking compared to China, Finland, or such homogenous countries is of little significance. These findings have global implications and political ramifications, which are not limited to Israel's individual case. As illustrated in this paper, trends of improvement in academic performance reflected in measures of self-change may not match the relative level of performance itself. In line with the position of other researchers (Bonnet, 2002; Feniger et al., 2012, Goldstien, 2004; Husen, 1987; Romainville, 2002), these findings come to warn against the problem of comparing the academic performance of students from different cultural backgrounds and other socio-economic variables. Overly focusing on the countries' average achievements as a measure of international educational comparison may create a false impression regarding the national educational level, instil public panic, and even mislead the educational policy.

REFERENCES

Anderson, J. O., Lin, H. S., Treagust, D. F., Ross, S. P., & Yore, L. D. (2007). Using large-scale assessment datasets for research in Science and Mathematics education: Programme for International Student Assessment (PISA). *International Journal of Science and Mathematics Education*, *5*(4), 591–614.

Atkin, J. M., & Black, P. (1997). Policy perils of international comparisons: The TIMSS case. *Phi Delta Kappan*, *79*(1), 22.

Barrett, G. V., & Alexander, R. A. (1989). Rejoinder to Austin, Humphreys, and Hulin: Critical reanalysis of Barrett, Caldwell, and Alexander. *Personnel Psychology, 42*(3), 597-612.

Ben-David, D. (2003). Israel's educational system in a socio-economic perspective in the age of globalization. *Economic Quarterly, 50*(1), 47–71. [in Hebrew].

Ben-David, D. (2010). "Israel's education system: An international perspective and recommendations for reform," in Dan Ben-David (ed.), *State of the Nation Report: Society, Economy and Policy in Israel 2009* (pp. 115–156). Israel: Taub Center for Social Policy Studies.

Blake, J. (1989). "Number of siblings and educational attainment." *Science, 245,* 32–36.

Blass, N. (2016). International exams and their importance to Israel's education system. Jerusalem: Taub center.

Bock, R. D. (1976). Basic issues in the measurement of change. *Advances in psychological and educational measurement,* 75–96.

Bonnet, G. (2002). Reflections in a critical eye: On the pitfalls of international assessment. *Assessment in Education: Principles, Policy & Practice, 9*(3), 387–399.

Coleman, J. S. (1975). Methods and results in the IEA studies of effect of school on learning. *Review of Educational Research, 45,* 335–386.

Committee for Measurement and Evaluation in Education (2005). *Proposal for a framework for curricula and professional development: Measurement and evaluation in education*. [in Hebrew].

Feniger, Y., & Lefstein, A. (2014). How not to reason with PISA data: An ironic investigation. *Journal of Education Policy, 29*(6), 845–855.

Feniger, Y., Livneh, I., & Yogev, A. (2012). Globalization and the politics of international tests: The case of Israel. *Comparative Education, 48*(3), 323–335.

Feniger, Y., & Shavit, Y. (2011). *Fertility and educational achievement: Israel in comparative perspective.* (Working papers, 5 October). Israel: Taub Centre.

Goldstein, H. (2004). International comparisons of student attainment: Some issues arising from the PISA study. *Assessment in education: Principles, policy & practice, 11*(3), 319–330.

Husén, T. (1987). Policy impact of IEA research. *Comparative Education Review, 31*(1), 29–46.

National Research Council. (2003). *Assessment in support of instruction and learning: Bridging the gap between large-scale and classroom assessment.* (Workshop report). National Academies Press.

OECD (2003). *PISA 2000.* Available from http://www.oecd.org/edu/school/programmeforinternationalstudentassessmentpisa/33690591.pdf

OECD (2005). *PISA 2003 Technical Report*. Available from https://www.oecd.org/edu/school/programmeforinternationalstudentassessmentpisa/35188570.pdf

OECD (2007). *PISA 2006: OEI*. Available from http://www.oei.es/historico/evaluacioneducativa/ResumenEjecutivoFinalingles.pdf.

OECD (2010). *PISA 2009 results: What students know and can do*. Retrieved from http://www.oecd.org/pisa/pisaproducts/48852548.pdf

OECD (2014). *PISA 2012 results in focus: What 15-year-olds know and what they can do with what they know*. Retrieved from https://www.oecd.org/pisa/keyfindings/pisa-2012-results-overview.pdf.

OECD *(2016). PISA 2015: Results in Focus.* Retrieved from https://www.oecd.org/pisa/pisa-2015-results-in-focus.pdf.

Rochex, J. Y. (2006). Social, methodological, and theoretical issues regarding assessment: Lessons from a secondary analysis of PISA 2000 literacy tests. *Review of Research in Education*, *30*(1), 163–212.

Romainville, M. (2002). Du bon usage de PISA. *Revue Nouvelle*, *115*(3–4), 86–99.

Sadler, T. D., & Zeidler, D. L. (2009). Scientific literacy, PISA, and socioscientific discourse: Assessment for progressive aims of science education. *Journal of Research in Science Teaching*, *46*(8), 909–921.

Sjøberg, S. (2007). PISA and "Real Life Challenges": Mission Impossible? In S. Hopman, (Ed.), *PISA according to PISA. Does PISA keep its promises?* Wien: LIT Verlag.

State of Israel, National Task Force for the Advancement of Education in Israel (Dovrat Committee) (2004). *National Education Program. Part A: The main recommendations*. Jerusalem: Hamehaber [in Hebrew].

Turner, R., & Adams, R. J. (2007). The programme for international student assessment: An overview. *Journal of applied measurement*, *8*(3), 237.

Willett, J. B. (1989). Questions and answers in the measurement of change. In Ernest Z. Rothkopf (Ed.), *Review of research in education* (Vol 15, pp. 345–22). Washington, D.C.: American Education Research Association.

Willett, J. B. (1994). Measurement of change. In T. Husen and T. N. Postlethwaite (Eds.), *International Encyclopedia of Education* (2nd ed, pp. 671–678). Oxford, UK: Elsevier Science Press.

Wuttke, J. (2007). Uncertainties and bias in PISA? In S. Hopman, (Ed.), *PISA according to PISA. Does PISA keep its promises?* Wien: LIT Verlag.

Yogev, A., Livneh, A., Feniger, Y. (2009). Singapore instead of Karkur? International achievement tests and the globalization of educational goals. *Megamot, 46*(3), 337–355. [in Hebrew].