Conceptualizing a Multifaceted Approach of Professional Evaluation Programs and Educational Outcomes

Tyrone Bynoe, University of the Cumberlands, KY, USA
Iris Ben-David, Bar-Ilan University, Israel

The North American Conference on Education 2014
Official Conference Proceedings

Abstract
Educational outcomes are the infrastructure for the future state competitiveness and cohesiveness. Evaluation has a strong effect on educational outcomes. This paper develops a conceptual framework for a more effective method to evaluate improvement in educational outcomes. The method we develop is based on adjusting Chernoff faces (Chernoff, 1973) to the educational arena. We display multivariate educational outcomes in the shape of a human face. The individual parts, such as eyes, ears, mouth and nose represent values of the variables by their shape, size, placement and orientation. At the preliminary step, we asked 50 Israeli educationalist and educators to indicate the most important educational outcomes (for their opinion). At the second step, we proposed a list of facets (that were generated at the preliminary step) and asked 200 Israeli teachers and experts to rank them from 1 (most important) to 10 (less important). The highly ranked facets were included in our model. The proposed model of evaluation suggests an alternative concept for educational improvement. States that will continue on focusing on a one facet linear approach might not be able to keep their relative advantage in the rapid changed global society. The focus on achievements and gain in achievements might be very costly having a tradeoff between achievements' gain and creativity.

iafor
The International Academic Forum
www.iafor.org

1. Introduction

The goal of this research is to develop a multifaceted conceptualization of improvement in Educational Outcomes. This study is significant because a state's accountability system governs schools outcomes and thereby affects the state's ability to compete in global era and state cohesiveness. The core assumption of this research project is that educational outcomes are multifaceted, whereas the current accountability systems are uni-faceted. They focus, for the most part, on standardized tests, which is only one facet of schooling, albeit important. They thus neglect the multifaceted nature of schooling (e.g., values, citizenship, solidarity), in turn contributing to a less attuned education system, which in turn diminish states' competitiveness and social cohesiveness. Therefore, we developed an alternative, multifaceted conceptualization of improvement in educational outcomes that can affect accountability policy. We claim that such an alternative approach to improvement in educational outcomes contributes to state's competitiveness and social cohesiveness.

In Section 2, we explore the current approaches for evaluation of improvement in Educational Outcomes most states use from a comparative and international perspective. In Section 3, we review the relationships between education, state competitiveness, and cohesiveness. Section 4 proposes a new approach for evaluating improvement in Educational Outcomes. In Section 5, we discuss the extent and the way in which the study's suggested, multifaceted conceptualization might influence state competitiveness and cohesiveness.

The questions addressed in this paper are as follows:

1. What characterizes the current approach of evaluating improvement in Educational Outcomes?
2. What is the potential effect of a multifaceted approach to evaluate improvement in Educational Outcomes? How would it affect state competitive ability and state cohesiveness?
3. What should be the facets in the multifaceted approach?

2. Literature Review

The following subsection introduces the literature on evaluating improvement in Educational Outcomes with respect to growth models, and this review explains the trends of how growth models distinguished from status models. The commonwealth of Kentucky will be consistenly referred to in this review as an example to how both student performance models emerged, featuring how the growth model eventually succeeded the status model.

Not until the age of accountability did most models of student performance analyze outcomes per se. In response to the Effective Schools' literature authored originally by Ron Edmonds and the debate between the James Coleman and Henry Levin regarding whether additional school funding mattered to improve student achievement, most analysis was on the inputs or ingredients that were believed to explain student effectiveness. This analysis never determined the measures or predictors of student outcomes that explained student achievement or school

effectiveness. The widespread state Supreme Court case in Kentucky in 1989 overturned over 700 pieces of legislation on education policy and instituted a state-wide outcomes-based education program which was to be systematically assessed. As Kentucky responded to A Nation At Risk and ensuing federal government initiatives to institute accountability, other state education agencies – from a couple of years earlier to several years after – proceeded to reform their systems, especially in response to America 2000 (1991), Goals 2000 (1994) and No Child Left Behind (NCLB - January 2002). By the time NCLB reauthorized the Elementary and Secondary Act of 1965, the federal government's call for accountability -- through higher curricular goals, testing of these goals, analysis of disaggregated test data by subgroup, and expected student performance based on annual yearly targets -- required together a systematic assessment of Educational Outcomes. Stated another way, with the guidance now of the National Assessment for Education Progress, the prospects for assessing systematically Educational Outcomes began based on essentially an outcome-based curricular program.

To understand how the emphasis of assessing Educational Outcomes emerged during the beginning of national accountability policy, an analysis of monitoring systems assessing Educational Outcomes – and not school, leadership or institutional inputs – necessitates a discussion about the changes in practitioner supervision to evaluation typifying both the Static Model and the Growth Model during this period beginning with the clinical model of teacher supervision.

The clinical supervision model from the 1960s to the 1980s typified a joint effort of teachers and administrators to improve teaching quality. Whether implemented by Ronald Goldhammer or Madeline Hunter, clinical supervision contained at least three elements, featuring a pre-conference, observed teaching, and a post-conference. The goal behind clinical supervision was to produce master teachers. During the 1980s when the clinical model was widely used*, A Nation At Risk* was officially issued in 1983, characterized the mediocrity of public education as a national security threat, and called for widespread reform in all state education systems. As curricular reform movements began during the mid-1980s to late 1980s, a static model of student achievement was gradually introduced (Marzano, Frontier, & Livingston, 2011).

Meanwhile, the mid-1980s witnessed the emergence of developmental/reflective models in teacher supervision that replaced the clinical model. Two key elements characterized this period of teacher supervision stressing developmental/reflective models. The first was a form of differentiating supervision between microscopic evaluation for probationary teachers and macroscopic evaluation for tenure teachers. Microscopic teacher evaluation required the supervisor to provide a thorough and prescriptive evaluation of the probationary faculty to ensure that these teachers matured in the prerequisite areas of sound pedagogy. In contrast, macroscopic evaluation contained a generic assessment and affirmation of a proven, tenured teacher. Nonetheless, effective administrators did not allow the tenure status to preclude the need to conduct a microscopic and detailed evaluation on an ineffective tenured teacher, especially if this meant the need to aggregate evidence for inevitable disciplinary proceedings. The goal behind the differentiated approach to teacher supervision was to improve the school's teacher quality.

The reflective model of teacher evaluation endeavored to supply direct support for teachers through professional development programs, and supervisors instructed teachers to use action-research to rethink ways of enabling teachers to improve intentionally their instructional effectiveness. As teacher quality began to improve through professional development and as teachers became more intentional in their practice from reflective thinking through action-research, practitioners emphasized that the imperative of supervision was to improve student achievement. This emphasis of improving student achievement as an outcome of teacher supervision was expected alongside developing state school reform programs of the 1980s and 1990s. Once gradual performance goals were instituted in these programs, teacher supervision switched to evaluation, and the static model of student performance characterized this evaluation. (Marzano, Frontier, & Livingston, 2011).

As static models were used in emerging state accountability programs during the 1990s and national accountability with NCLB, growth models were only used in specific schools as methodologies of program evaluation, such as in outcome-based or impact evaluations. Nonetheless, several factors would call for accountability systems to consider using growth models in their evaluation programs due to the perceived shortcomings of then existing static models of evaluation. Along with inefficacy of clinical supervision to address the policy-makers' mission to improve student achievement, one major concern was the uncertainty that single performance measures in accountability policies could provide valid and reliable of measures to evaluate student achievement. In response to this concern, policy makers reasoned that schools being complex institutions could not be evaluated by a single target measure. These policy-makers called for multiple-measures of student achievement as well as a 360 degree feedback-loop of both quantitative and qualitative measures to assess school performance more holistically.

A second concern was the assumption regarding the attainment of linear growth that schools sought after when using a static model of evaluation (Elmore, 2007, p. 1). This proved significantly true in Kentucky, which encountered three iterations of state instructional programming: first with performance-based testing in 1990; second with standardized testing in 1994; and third with completely revised standardized testing in 2010. A standardized assessment program called Comprehensive Accountability Test System (CATS) replaced a performance-based assessment program known as Kentucky Instructional Results Information System (KIRIS) in 1994. The Kentucky Department of Education (KDE), then, instituted an accountability index of school performance reinforcing the static model of student and school evaluation. From 1994 to 2014, schools were expected to improve ten percentage points every two years and reach the index of 100 out of 140, which measured an attainment of *proficiency* (Whitford & Jones, 2000, 9-23). But by 2007, it was clear that 63 percent of all Kentucky public schools were not improving every two years with gains of at least ten percentage points. In fact, these schools were not on target to reach the accountability index of 100 out 140 by 2014.

Not only did schools demonstrate fluctuations in meeting the index target of this static model, but many schools digressed, displaying downward-sloping student performance decline (Council for Better Education; Perkins & Sexton, 2009, pp. 23-29). The fluctuations and downward-sloping movement in school performance compelled KDE to terminate the CATS and the static model of the accountability

index, particularly since its assessments were not directly aligned to NCLB assessments (Innes, 2005). Simply stated, Kentucky's assessment program of static growth proved unreasonable to actualize. This explains why in its third iteration of instructional programming, Kentucky became the first state to adopt the Common-Core curriculum during February, 2010 under a state education reform initiative from its Senate Bill 1 known as Unbridled Learning and to revise completely its accountability system featuring a battery of new assessments and a different practitioner evaluation system which adopted several aspects of a growth model (SB1 09RS, 2009).

The federal government's Race-to-the-Top initiative also caused states to rethink their static models of teacher evaluation. Through Race-to-the-Top, the federal government engaged states in a national competition to improve their accountability programs against formidable school reform guidelines as an incentive to be awarded extra federal monies. This prompted states to revisit the static model in their evaluation systems, and persuaded these states to consider various growth models in their accountability systems (Campbell, 2013, p. 43).

Meanwhile, the Obama administration encountered resistance in Congress to rewrite NCLB. This resistance mirrored the intransigence that members of Congress exhibited in passing promptly the federal budget, which lead to Congress implementing a sequestration with the goal of recapturing fiscal constraint legislated in the Gramm-Rudman-Hollings balanced budget act of 1985. Given Congressional intransigence and stalling to rewrite NCLB, the Obama administration gave states the opportunity to obtain waivers to comply with NCLB provided that states developed and submitted innovative accountability plans that the U.S. Department of Education endorsed (Duncan, June 2013). Subsequently, states *fine-tuned* their accountability policies, scrapped static models from their practitioner evaluation programs, and adopted growth models and growth indicators in these programs.

Concurrent innovations continued in teacher supervision as states revised their accountability programs in response to federal government influence and state-by-state adoption of the Common-Core curriculum. Tucker and Strange made one important innovation in teacher supervision when they argued that effective supervision requires the input of both student gain scores and growth scores to provide valid and reliable feedback on teacher effectiveness. In *Linking Teacher Evaluation and Student Learning*, Tucker and Stronge insisted that student growth scores constituted more valid and reliable measures of student performance, argued that other feedback sources assess teachers more effectively than observations alone, and called for various student performance scores and observations to typify teacher evaluations (Tucker & Stronge, 2005). Simply stated, Tucker and Strange established a paradigm shift in teacher supervision to be replaced with program evaluation. As evaluation is to replace supervision, Tucker and Strange argued for the growth model to replace the static model (Marzano, Frontier, T. & Livingston, D., 2011).

When reflecting on how supervision developed in response to the rise of accountability and how evaluation eclipsed supervision when state accountability policy was revised, the underlying theme behind this paradigm shift in the assessment of teachers and school practitioners has been the compelling need to improve the teaching profession. In a nutshell, growth models replaced static models due to a

systemic need to improve the quality of teaching since this profession has not been rated with the quality of integrity and effectiveness of other professions: ie. law, medicine, engineering, etc. This shift does not necessarily prove that growth models constitute ideal accountability programs or that they really improve the profession of teaching. Nevertheless, the review's explanation of this shift justifies the researchers' curiosity to ask questions related to the *real and enduring worth* of growth models in accountability programs.

3. Improvement in Educational Outcomes

In recent years there has been an increment in the quality of performance-related data available to schools to inform school improvement. Yet there remains scope for more refined and intelligent measures that will better indicate how schools are progressing in improving the learning outcomes of students.

Meyer (1997) claims that the indicators commonly used to assess school performance-average and median test scores are highly flawed. They tend to be contaminated by student mobility and by non-school factors that contribute to student achievement (e.g. student, family and community characteristics and prior achievement). Meyer and Dokumaci (2011) assert that the conceptually-appropriate indicator of school performance is the value-added indicator. The value-added indicator measures school performance using a statistical model that includes, to the extent possible, all the non-school factors that contribute to growth in student achievement. The objective is to statistically isolate the contribution of schools to student achievement growth from these other factors.

3.1 Growth models

Growth models generally refer to models that measure progress by tracking the achievement scores of the *same* students from one period to the next with the intent of determining whether or not, on average, the students made progress. Growth models assume that student performance, and by extension school performance, is not simply a matter of where the school is at any single point in time, and that a school's ability to facilitate academic progress is a better indicator of its performance. Growth models can vary, but in general, they account for the potentially negative spurious relationship between status and growth, for the effect of status on growth, and for the effect of student inputs on growth. The greater the number of occasions (years) used to estimate growth, the less initial performance will be related to growth (Goldschmidt et al., 2005)—this means growth will be less and less related to indicators of school performance that are based on cross-sectional indicators. In general, we would expect all students to demonstrate some academic progress across grades, but some schools will still exhibit more growth than others, on average.

3.2 Value-added models

Value added models are one type of growth model in which student background characteristics and/or prior achievement and other data are used as statistical controls in order to isolate the specific effects of a particular school. Value-added approaches aim to provide a clearer indication of the contribution a school makes to the progress of its students by adjusting for the impact of non-school influences on student

performance. Value-added modeling (VAM) can also be used to create projections of school performance that can assist in planning, resource allocation and decision making (OECD, 2008). Value-added measures have emerged internationally as a means of assessing school performance. The value-added approach recognizes that students have different levels of capability and come from different environments, and that these factors will influence each student's rate of educational progress.

The main purpose of VAM is to separate the effects of non-school-related factors (such as family, peer, and individual influence) from a school's performance at any point in time so that student performance can be attributed appropriately. A value-added estimate for a school is simply the difference between its actual growth and its expected growth. It is important to note that schools can demonstrate positive achievement growth, but still have a value-added estimate that is negative (i.e., the school demonstrated growth, just not as much as we would have predicted given the student inputs available to the school).

The term value-added was initially popularized as part of the Tennessee Value-Added Accountability System (TVAAS) (Sanders, Saxton, & Horn, 1997; Ballou, Sanders, & Wright, 2004; McCaffrey, Lockwood, Koretz, Louis, & Hamilton, 2004; Ballou, 2005). Hayes and Taylor (1996) using Dallas school data, found that the schools' value-added explains 10 percent of the total explained variation in student performance.

Value-added modeling is most common in the U.S. and the United Kingdom. In the U.S., value-added modeling has recently been applied also to measuring individual teachers' contribution to student learning (Loeb et al., 2007; Boyd et al., 2006), revealing that teachers play an important role in this respect (Loeb et al. 2007). In the United Kingdom, value-added modeling is used (1) in Performance Tables, which provide information to parents and hold schools to account; (2) in systems for school improvement, where the data are used for self-evaluation and target setting; (3) to inform school inspections, which are now tied into the school improvement process; (4) to help select schools for particular initiatives; and (5) to provide information on the effectiveness of particular types of school or policy initiatives (Ray, 2006).

However, using VAM for policy initiative needs to be treated with delicacy. Briggs, Weeks, & Wiley (2008), draw attention to the pitfalls of using VAM for policy initiative. They find that the precision of value-added estimates can be quite sensitive to the combinations of choices made in the creation of the scale. They conclude that when VAM are being used for the purposes of high-stakes accountability decisions, its sensitivity is most likely to be problematic.

Nonetheless, this method is also in use in other parts of the world. Smaller regional and pilot initiatives have also been developed in a number of countries. OECD member countries were invited to join the project in July 2006. Thirteen countries chose to participate in the project: Australia; Belgium (Flemish Community); Czech Republic; Denmark; France; Netherlands; Norway; Poland; Portugal; Slovenia; Spain; Sweden; and the United Kingdom (OECD, 2008).

4. Promoting state competitiveness and social cohesiveness

4.1 Education and state competitiveness

The increasing need for state competitiveness in the global market is due to the accelerating processes of globalization, in particular the challenge to many states to sustain their position in the market relative to other states (Green, Mostafa, & Preston, 2010).

The literature linking education and competitiveness views education as an infrastructure for advancing state competitiveness. Reiljan, Hinrikus, and Ivanov (2000) argue that the ability to achieve competitiveness is more important than competitiveness itself, because it guarantees recuperation if competitiveness is lost for some reason. The importance of education accumulated in human capital development is highlighted in the light of this argument. Furthermore, they claim, one important aspect that should be evaluated to predict a country's future competitiveness is education. Their model concludes that an individual's competitiveness is mainly a derivative of his or her education, whereas the competitiveness of a state depends much upon the ability of a nation to create an environment that favors education for development.

Both primary and secondary education significantly contribute to economic development and growth. The literature recognizes human capital development and demonstrates how increased investment in education provides future returns to the economy through increases in labor productivity (Hanushek & Kimko, 2000; Krueger & Lindahl, 2000). Moreover, better quality education increases average earnings and productivity and reduces the likelihood of social problems that, in turn, are harmful for economic development.

Sahlberg (2006) claims that successful economies compete on the basis of high human capital development, which is best guaranteed by educated personnel. He argues that globalization has increased economic competition between countries. Furthermore, Sahlberg highlights the general assumption that, to increase competitiveness, citizens must acquire knowledge, skills and attitudes necessary for civic success and the knowledge-based economy. He concludes that the key features of education reform policies compatible with competitiveness are those that encourage flexibility in education systems and creativity in schools.

4.2 Education and social cohesiveness

A salient argument in the literature linking education and social cohesion is that the distribution of education attainment affects social cohesion. Thus, countries with education systems producing more equal outcomes are likelier to promote future social cohesion than countries in which education is distributed less equitably (Green & Preston, 2001).

Beauvais and Jenson's (2002) review of the literature concerning education and social cohesiveness also indicates that state education is an important ingredient for fostering social cohesion. Moreover, a state's economic and social policies (for example, its investment in children through education) are an important factor for

achieving future cohesion. Additionally, this review points out that UNESCO also argues for the importance of education and education policy for social cohesion. Beauvais and Jenson conclude, therefore, that if globalization produces greater demographic diversity, then public policy can be used to improve social cohesion.

5. A new approach to evaluate Improvement in Educational Outcomes

We argue that focusing on a one (or two) facet to evaluate improvement in the educational process (e.g., students' academic achievements, and instructional practices) is sometimes narrow as the educational process is very complex. One may make an impressive gain in academic achievement and yet "pay the price" (within the tradeoff view) in terms of lower social engagement, or lower level of values. Thus, one may exhibit low gain in academic performance yet lead in innovation and creativity. Focusing on the gain in students' performance and on the development in teachers' instructional practices neglects many issues that educational systems account for.

Table 1 presents the facets and their (averaged) ranking as we generated from the preliminary step of our research. The following table is comprised of facts that we collected from educators and educationalist that were interviewed. The questions were as follows: What facets do educational system accounts for? Please indicate which of the facets (among the facets you have mentioned) is perceived (by you) as the most important and please rank them (the rank 1 is given to the most important facet).

The following Table 1 presents the averaged ranks.

Table 1: Facets and their averaged ranks

Educationalist		Educators				
Facet	(1) (n=20)	(2) Math (n=30)	(3) Science (n=20)	(4) Language (n=30)	(5) Art & Sport (n=20)	(6) Averaged rank	**(7) Final aver ranks**
Academic Achievements	1	1	3	10	1	3.58	2
Social engagement	3	4	13	8	3	6.17	5
Values	2	8	1	1	2	3.08	1
Happiness	12	6	12	6	12	9.00	10
Leadership	5	14	7	13	5	9.58	11
Health (hunger)	6	5	11	2	6	5.58	3.5
Optimism	4	13	10	7	4	8.00	7
Self-Efficacy	11	2	4	3	11	5.58	3.5
Self- Awareness	7	3	14	14	7	8.92	9
The volition to succeed	8	9	2	4	8	6.25	6
Innovation	9	13	8	11	9	10.33	12
Creativity	10	12	9	12	10	10.83	14
Violence	13	7	5	5	13	8.17	8
Instructional practices	14	10	6	9	14	10.42	13

A total of 120 Israeli educators and educationalist filled in a questionnaire (100, and 20, respectively). They were asked to rank from the most important (1) to the less important (14) facets of education that were collected at the preliminary step of interviews (Columns 2 to 5). They could also suggest other facets to be included or suggest omitting some of the proposed facets. The teachers' seniority is averaged (10 to15 years of seniority). Most of the teachers that reply are holding bachelor degree and teacher's certificate. We asked 20 more Professors of education to rank the proposed facets (Column 1).

The averaged response of each group is presented in Table 1 (Columns1 to 5). The final step was to calculate weighted average considering for each group size and to rank these averaged ranks (Column 6) into final ranks (Column 7). To exemplify, the lowest averaged rank 3.08 was assigned a final rank of 1, and so forth. Additional considerations were enacted upon equal averaged ranks. In this case the averaged final ranks were assigned to each of these facets (e.g., the facets Health/Hunger and Self-Efficacy both averaged rank was 5.58 account for the third and the forth ranks, therefore, their final assigned rank is 3.5 and the following rank is 5).

The results from the questionnaires indicates that Values are perceived as the most important facet of education (was assigned a final averaged rank of 1, Column 7). It was also found that students' achievement is also perceived as a very important facet of education. Finally, the facet Instructional practices was perceived as less important ant was assigned the rank 13 (out of 14). We moved forward to present the different facets using Chernoff Faces.

Table 2: The components of Chernoff face and their respective education improvement facets

Element in Figure 1	Facet/ time
1. Size of face	**Academic Achievements**
2. Forehead/jaw relative arc length	Constant
3. Shape of forehead	**Social engagement**
4. Shape of jaw	**Values**
5. Width between eyes	**Instructional practices**
6. Vertical position of eyes	**Leadership**
7. Height of eyes	**Health (hunger)**
8. Width of eyes	**Creativity**
9. Angle of eyes	**Self-Efficacy**
10. Vertical position of eyebrows	**Self-Awareness**
11. Width of eyebrows (relative to eyes)	**The volition to succeed**
12. Angle of eyebrows (relative to eyes)	**Innovation**
13. Direction of pupils	Constant
14. Length of nose	**Violence**
15. vertical position of mouth	Constant
16. Shape of mouth	**Happiness**
17. Mouth arc length	**Optimism**

Table 2 presents the facets and their graphic presentation. To exemplify, the academic achievement of a student, school or a state is presented by the size of the face. Large face means high academic achievement and visa verse. Improvement in academic

achievement is represented by enlarging size of the face. Happiness is represented by the shape of the mouth and so forth.

The relative representation of Cernoff faces is useful in education because education is often considered as a positional good. Excellent performance of one worsens the relative position of the other. Specifically, improvement in education is relative. This point can also be addressed from the point of view of the state. The rate of improvement of one state is dependent on the rate of improvement of other states. To this end, Figure 1 illustrates 10 faces each represent an alternative state (the first and the last faces, 1 and 12, are only sets as reference point where the first face, 1 represents the worst hypothetical prototype and the last face, 12 represents the best).

Moreover, since improvement in education is not linear, the Chernoff faces are representing improvement better compared with VAMs as they do not assume linearity.

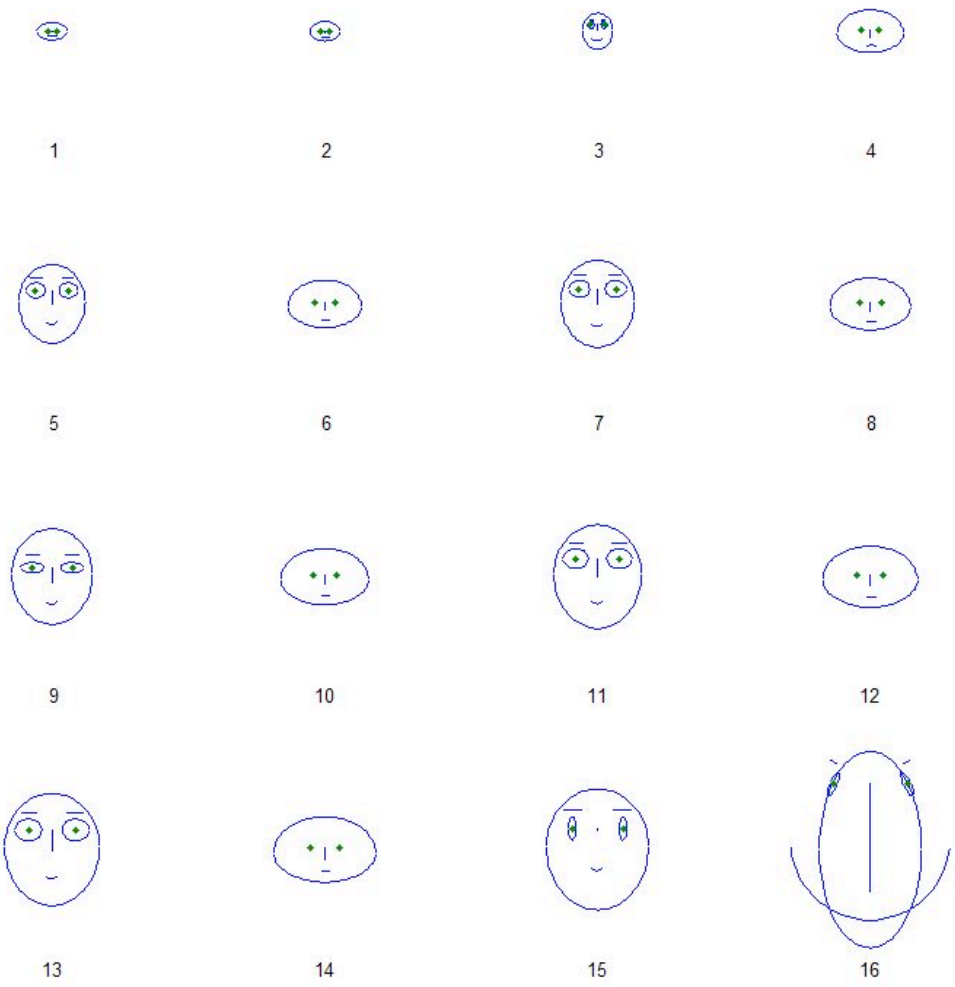

Figure 1: Face charts for students improvement along time

Figure 1 presents a multivariate representation of educational improvement. It shows an example of Chernoff faces of one student along time (or of one school/district/ state) using MATLAB statistical toolbox. It can also be used to represent a comparative view on different states.

The argument of this paper is twofold: (a) Education is multifaceted, and therefore, a one facet approach used to evaluate or to measure educational performance is too narrow. (b) Educational improvement is relative and not linear, and therefore, using VAMs to measure improvement in a linear manner is not effective. Chernoff faces abele as to take a multivariate approach towards education and to represent a nonlinear improvement. The 14 "faces" presented in Figure 1 (numbered 2 -15) represent different prototypes of students and their educational improvement.

To exemplify, we further analyze "face" number 14 and "face" number 15. Table 3 summarizes the characteristics of both these prototypes used in our example.

Table 3: Characteristics of two prototypes on a unified scale (0=lowest, and 100=highest)

Facet	Prototype 14	Prototype 15
Academic Achievements	100	100
Social engagement	20	80
Values	0	100
	0	100
Leadership	0	100
Health (hunger)	0	100
Creativity	0	100
Self-Efficacy	0	100
Self-Awareness	0	100
The volition to succeed	0	100
Innovation	0	100
Violence	100	0
Happiness	0	100
Optimism	0	100

Both prototypes resemble in their high academic achievement yet they are opposite to the extreme in all other facets. Specifically, all other characteristics of prototype 15 (e.g., prototype 15 is extremely innovative and happy and not at all defined as a violent student) are high compared with the characteristics of prototype 14 (e.g., prototype 14 is extremely violent and not at all defined as innovative or happy). If we were interested solely in measuring and evaluating teachers effectiveness based on the improvement gained in their students' academic achievement than we would have mistakenly rewarded both prototype 14 and 15. However, the Charnoff faces presented in Figure 1 lead very smoothly to the conclusion that prototype 15 should be rewarded or at least considered more effective. (e.g., large face, smiling mouth, and the tiny nose account for high academic achievement, happiness, and non-violent, respectively).

The above mentioned example uses solely one pair of "faces" however, Figure 1 illustrates 6 more pairs (faces 2 and 3, faces 4 and 5, etc.) each pair differ solely in the facet of academic achievement (i.e., the level of academic achievement illustrated in

faces 2 and 3 is the lowest-0, the level of academic achievement illustrated in faces 4 and 5 is higher-50, the level of academic achievement illustrated in faces 6 and 7-60,… the level of academic achievement illustrated in faces 12 and 13 is 90, and so on) and the rest of the facets resembles our previous example. Again, if we would reward schools based solely on academic achievement prototype 2 and 3 were not differentiated as both represent the lowest level of achievement. Yet, prototype 3 is far more improved compared with prototype 2 in all other facets as illustrated in Figure 1.

6. Discussion

The method by which educational outcomes is evaluated has a far going effect on the process of learning and teaching. It also has a tremendous effect on the state ability to compete globally. Most western states acknowledged education as a multifaceted process as defined by the long list of educational goals (and objectives). Yet, the current method of evaluation focuses on a narrow approach. Many states (see Appendix Table 1) are developing models of evaluation that are focused on two facets of education thus neglect other goals of education.

One can argue that many educational outcomes cannot be measured. Yet, there is a growing consensus amongst educators and educationalist that perceived education as has more than two facets that can be measured.

Others might argue that the goals of education set by policy makers are only rhetoric and therefore focusing solely on measuring and evaluating academic achievement is more than enough as demonstrated by the Israeli case. The reform taken place in Israel recently has put the questions of this paper to the forth. While the U.S. approach is pro assessing sometimes to the extreme, the Israeli approach is very far from that. Up until recently the Israeli system used to evaluate schools using student's performance once at each schooling level (i.e., primary, lower secondary, and upper secondary school levels). After Supreme Court decision demanding transparency of the evaluation reports (both at the primary and lower secondary school levels), policy was reformed. Currently, the national evaluation is taking place once at the upper secondary level.

It is important to acknowledge that the Israeli law of education encompasses many goals (e.g., the goal of education is to encourage initiative and creativity). Yet, the evaluation of educational outcome in Israel is focused on students' academic achievement.

The U.S. law of education also encompasses many goals. Yet, similar to Israel the focus of its evaluation method is on students' achievement and (sometimes) on instructional practices of its educators, neglecting other important goals. In contrast with Israel the focus in the US is on the gain in achievement. However, the essence of education as multifaceted, though addressed in the law, is not translated into practice.

Furthermore, similar to the U.S. and to Israel the leading actors in the education evaluation arena (e.g., PISA, TIMSS) enact the same approach while reporting on educational systems globally, and measuring the improvement they gained.

To conclude, there is a need to change the perspective on educational improvement. States that will continue on focusing on a one facet linear approach might not be able to keep their relative advantage in the rapid changed global society. The focus on achievements and gain in achievements might be very costly having a tradeoff between achievements' gain and creativity.

We hope the nation state policy makers and global actors that shape the future of our societies will reform the policy of evaluating educational improvement to comply with the recommendations of this work.

References

Ballou, D. (2005). Value added assessment: Lessons from Tennessee. In Lissitz, R. W. (Ed.). *Value added models in education theory and application* (pp. 272-303). JAM Press: Maple Grove, Minnesota.

Ballou, D., Sanders, W., and Wright, P. (2004). Controlling for student back-ground in value-added assessment of teachers. *Journal of Educational and Behavioral Statistics,* 29(1), 37- 65.

Beauvais, C., and Jenson, J. (2002). *Social cohesion: updating the state of the research investigating the concept of social cohesion*. Canada: Canadian Policy Research Networks

Boyd, D., Grossman, P., Lankford, H., Loeb, S., and Wyckoff, J. (2006). How changes in entry requirements alter the teacher workforce and affect student achievement. *Education Finance and Policy*, 1(2), 176- 216.

Campbell, T. (January 2013). *Teacher supervision and evaluation: a case study of administrators' and teachers' perceptions of mini observations*. (Doctoral Dissertation, Northeastern University).

Chernoff, Herman. 1973. The use of faces to represent points in k-dimensional space graphically. *Journal of the American Statistical Association*, 68 (342): 361-368.

Council for Better Education. *Progress, but not enough progress: A briefing paper from the Council for Better Education.* Retrieved from: http://www.kycbe.com/downloads/ProgressCBE.pdf

Duncan, A. (June, 2013). *Key policy letters from the education secretary and deputy secretary.* Washington, DC: US Department of Education. Retrieved from: http://www2.ed.gov/policy/elsec/guid/secletter/130618.html

Elmore, R. (June 2007). The road to school improvement: It's hard, it's bumpy, and it takes as long as it takes. *Harvard Education Letter*, *23*(3), 1-3. Retrieved from: http://hepg.org/hel-home/issues/23_3/helarticle/the-road-to-school-improvement_229

Goldschmidt, P., Roschewski, P., Choi, K., Auty, W., Hebbler, S., Blank, R., and Williams, A. (2005). Policymakers' Guide to Growth Models for School Accountability: How do Accountability Models Differ? *A paper commissioned by the CCSSO Accountability Systems and Reporting State Collaborative on Assessment and Student Standards*.

Green, A., Mostafa, T. and Preston, J. (2010). *The Chimera of Competitiveness: Varieties of Capitalism and the Economic Crisis*. Centre for Learning and Life Chances in Knowledge Economies and Societies at: http://www.llakes.org.uk.

Green, A. and Preston, J. (2001).Education and social cohesion: Re-centering the debate, *Peabody Journal of Education and Development*, 76 (3 and 4), pp. 247-284.

Hanushek, E., and Kimko, D. (2000). Schooling, labor-force, quality and the growth of nations. *American Economic Review*. 90(5), 1184–1208.

Hayes, K. J. and Taylor, L.L. (1996). Neighborhood school characteristics: What signals quality to homebuyers? *Federal Reserve Bank of Dallas Economic Review*, fourth quarter, 2-9.

Innes, R. (November 2005). *CATS: An Inadequate NCLB basis for school improvement.* Bowling Green, KY: The Bluegrass Institute. Retrieved from: http://www.freedomkentucky.org/images/b/b6/CATSInadequateBasisforNCLB.pdf

Kentucky Legislative Regulations. (2009) SB1 09RS -- Retrieved from: http://www.lrc.ky.gov/record/09RS/SB1.htm

Krueger, A., and Lindahl, M. (2000). Education for growth: Why and for whom? *Working Paper #7591.* Washington, DC: National Bureau of Economic Research.

Loeb, S., Boyd, D., Lankford, H., Rockoff, J., and Wyckoff, J. (2007). The narrowing gap in teacher qualifications and its implications for student achievement in high poverty. Working paper 10. Center for Analysis of Longitudinal Data in Education Research (CALDER).

Marzano, R., Frontier, T., & Livingston, D. (2011). *Effective supervision.* Alexandria, VA: ASCD.

McCaffrey, D. F., Lockwood, J., Koretz, D., Louis, T. A., and Hamilton, L. S. (2004). Models for value- added modeling of teacher effects. *Journal of Educational and Behavioral Statistics*, 29(1), 67- 101.

Meyer, R.H. (1997). Value-added indicators of school performance: A primer. *Economics of Education Review,* 16(3), 283-301.

Meyer, R.H. and Dokumaci, E. (2011). Value-added models and the next generation of assessments. Paper presented at the annual conference of the Association for Education Finance and Policy, Seattle Washington, 24-26 March.

Organisation for Economic Co-operation and Development (OECD). (2008). *Measuring Improvements in Learning Outcomes: Best Practices to Assess the Value-Added of Schools*. Paris: OECD.

Perkins, S. & Sexton, R. (December 2009). *Substantial and yet not sufficient: Kentucky's effort to build proficiency for each and every child*. Education, Equity and the Law, No. 2 The Campaign for Educational Equity: TC-Columbia University. Retrieved from: http://files.eric.ed.gov/fulltext/ED523999.pdf

Ray, A. (2006). School value added measures in England. Paper for the OECD project on the development of value-added models in education systems. UK: Department for education and skills.

Reiljan, J., Hinrikus, M., and Ivanov, A. (2000). Key Issues in Defining and Analysing the Competitiveness of a Country. University of Tartu Economics and Business Administration Working Paper No. 1.

Sahlberg, P. (2006). Education reform for raising economic competitiveness. *Journal of Educational Change*, 7(4), 259-287.

Sanders, W., Saxton, A., and Horn, S. (1997). The Tennessee value-added accountability system: A quantitative, outcome-based approach to educational assessment. In J. Millman (Ed.), *Grading teachers, grading schools* (pp. 137-162). Thousand Oaks, CA: Corwin Press.

Senate Bill 1 Highlights – 2009 Session of the Kentucky General Assembly. Retrieved from: http://education.ky.gov/comm/ul/documents/senate%20bill%201%20highlights.pdf

Stronge, J. & Grant, L. (2009). *Student Achievement Goal Setting: Using data to improve teaching and learning.* Larchmont, NY: Eye On Education.

Tucker, P. & Stronge, J. (April, 2005). *Linking teacher evaluation and student learning.*Alexandria, VA: ASCD.

Whitford, B. & Jones, K. (Eds.). (2000). *Accountability, assessment, and teacher commitment: Lessons from Kentucky's reform efforts.* Albany, NY: SUNY Press.

Contact email address: tyrone.bynoe@ucumberlands.edu