School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/nses20

Exploring whole school versus subject department improvement in Hong Kong secondary schools

James Ko[a], Philip Hallinger[a] & Allan Walker[a]
[a] Hong Kong Institute of Education, Tai Po, NT, Hong Kong
Published online: 11 Feb 2014.

To cite this article: James Ko, Philip Hallinger & Allan Walker , School Effectiveness and School Improvement (2014): Exploring whole school versus subject department improvement in Hong Kong secondary schools, School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice, DOI: 10.1080/09243453.2014.882848

To link to this article: http://dx.doi.org/10.1080/09243453.2014.882848

School Effectiveness and School Improvement, 2014
http://dx.doi.org/10.1080/09243453.2014.882848

Routledge
Taylor & Francis Group

Exploring whole school versus subject department improvement in Hong Kong secondary schools

James Ko, Philip Hallinger* and Allan Walker

Hong Kong Institute of Education, Tai Po, NT, Hong Kong

(*Received 5 March 2013; final version received 12 November 2013*)

Research on school improvement tends to assume that school improvement is a school-wide process. Nonetheless, some researchers have also proposed that secondary schools are comprised of subcultures centered on subject area departments. It has further been suggested that variations in the sociocultural organization of subject departments could produce differential results in the learning outcomes of students. To date, however, few empirical studies have examined the leverage offered by this perspective. The current study addresses this gap in the literature by analyzing changes in the learning outcomes of subject departments over a 3-year period as compared with patterns of whole school improvement in 47 Hong Kong secondary schools. The results support the proposition that theory, research, and practice on secondary schools will benefit from viewing their improvement from a perspective that takes into account the diversity of department-level conditions and learning outcomes.

Keywords: school improvement; education reform; secondary schools; Hong Kong; change; subject departments

Introduction

> Educators have become increasingly convinced that the characteristics of schools are important determinants of academic achievement. Since 1978, an extraordinary number and variety of programs have concentrated on a school effects interpretation of the relationship between student achievement and family background. Such programs represent a major education reform and derive from the fairly rapid educator acceptance of the research of Brookover and Lezotte (1979), Edmonds (1979), and Rutter, Maughan, Mortimore, and Ouston (1979) and a number of others who have studied the characteristics of effective and ineffective schools. (Edmonds, 1982, p. 4)

This quotation from Ron Edmonds in 1982 heralded the "modern era of school improvement". This era has been characterized by several common features:

(1) a belief in the efficacy of planned organizational change (Cuban, 2013; Firestone & Corbett, 1988; Fullan, 2007; Hall & Hord, 2002; Potter, Reynolds, & Chapman, 2002; Reynolds, 2010; Stringfield, Reynolds, & Schaffer, 2008a, 2008b);
(2) school improvement policy and practice informed by empirical research (Creemers & Kyriakides, 2008; Creemers & Scheerens, 1994; Edmonds, 1979, 1982; Hall, 2013; Hattie, 2009; Hawley & Rosenholtz, 1984; Murphy, 2013; Purkey & Smith, 1983; Reynolds, 2010; Rosenholtz, 1985; Stringfield, Reynolds, & Schaffer, 2012);

*Corresponding author. Email: Hallinger@gmail.com

(3) a research-informed conviction that schools have the capacity to ensure that all children are educated to common standard (Edmonds, 1979; Hattie, 2009; Hawley & Rosenholtz, 1984; Murphy, 2013; Potter et al., 2002; Reynolds, 2010; Schaffer, Reynolds, & Stringfield, 2012; Stoll & Fink, 1996); and
(4) focus on the whole school as the unit of analysis and target for change (Edmonds, 1982; Hallinger & Heck, 2011; Hawley & Rosenholtz, 1984; Huberman & Miles, 1984; Louis & Miles, 1990; Murphy, 2013; Slavin & Madden, 2006; Stoll & Fink, 1996).

These features are reflected in school improvement policies and practices as diverse as quality assurance, accountability frameworks, school development planning, data-based decision-making, standards-based curricula, and teacher evaluation. Indeed, it is no exaggeration to assert that these beliefs have, during the ensuing decades, become embedded in the modal approaches to school improvement employed by school systems throughout much of the world (Bryk, Sebring, & Allensworth, 2010; Creemers & Kyriakides, 2008; Mulford & Silins, 2003, 2009; Murphy, 2013; Reynolds, Stringfield, & Schaffer, 2006; Reynolds, Teddlie, Hopkins, & Stringfield, 2000; Stoll & Fink, 1996).

At the same time, however, a significant offshoot of this main vein of school improvement research has grown in a somewhat different direction. During the late 1980s and continuing to the present, researchers began to conceptualize and examine the role of subcultures associated with subject departments in secondary schools (e.g., Creemers & Kyriakides, 2008; McLaughlin & Talbert, 2001, 2006; Reynolds, 2010; Rosenholtz, 1989; Sammons, Thomas, & Mortimore, 1997; Schaffer et al., 2012; Scheerens, 1990; Scheerens & Creemers, 1989; Siskin, 1991, 1997; Stringfield et al., 2008a, 2008b). This line of inquiry has emphasized not only overall improvement of school as a whole, but also variability of department-level cultures and performance in secondary schools. Ainley (1994) captured the assumptions underlying this perspective on secondary school improvement:

> The nature of secondary schools raised additional issues because of their greater organizational complexity and because the outcomes of learning may involve a much wider range of areas of learning … in terms of research it is important to incorporate the department as a central component of high school organization. (pp. 14–15)

This line of research has yielded a strong conceptual and practical argument for adopting a more refined perspective on secondary school improvement. Yet, we observe that relatively few studies have actually analyzed secondary school improvement in terms of the variation in outcomes produced by subject departments in secondary schools (e.g., see also Sammons et al., 1997; Stringfield et al., 2008a). The current study addresses this issue by comparing patterns of school-wide improvement in a sample of secondary schools with improvement in their composite subject departments. The research aims to determine whether identifiable patterns can be established in the improvement of departments over time, and how these patterns compare to patterns of improvement in overall school performance. The study addresses the following research question: What is the relationship between change in whole school and subject department performance over time?

We note that this study of performance outcomes does *not* include an examination of "school inputs" or "explanatory factors" as is often the case in school improvement studies (e.g., Hallinger & Heck, 2011; Ko, Hallinger, & Walker, 2012; Mulford & Silins, 2009; Sammons et al., 1997; Sleegers, Geijsel, & Van den Berg, 2002). Instead, this research is

framed to offer insight into the relationship between patterns of change in subject departments and the school as a whole. We suggest that this delimited scope does, however, have interrelated implications for research, policy, and practice. In secondary schools, relying on a lens that focuses exclusively on "whole school performance" could lead researchers and practitioners to overlook important variability in both the capacity and performance of its key structural units (Sammons et al., 1997; Stringfield et al., 2008a, 2008b). From a practical perspective, this loss of information could prove critical to developing a clearer understanding of a secondary school's strengths and weaknesses. Gaining traction on this issue could inform the design of new change strategies. The study contributes to research on school improvement by providing a longitudinal, empirically grounded perspective on the analysis of change in learning outcomes in secondary schools.

Theoretical perspective

The literature on school improvement has tended to examine patterns of change in the educational processes and outcomes that characterize the school as an organizational unit (Murphy, 2013). Strongly influenced by studies of effective schools (e.g., Edmonds, 1979, 1982; Purkey & Smith, 1983), this approach begins with the assumption that the most powerful strategies to improve the quality of teaching and learning in schools are located in interventions that focus on the school as the key organizational unit (Crandall, Eiseman, & Louis, 1986; Creemers & Kyriakides, 2008; Firestone & Corbett, 1988; Geijsel, Sleegers, & Van den Berg, 1999; Geijsel, Sleegers, Van den Berg, & Kelchtermans, 2001; Hall, 2013; Hallinger & Heck, 2010; Jackson, 2000; Ko et al., 2012; Mulford & Silins, 2003, 2009; Murphy, 2013; Reynolds, 2010; Scheerens, 1990; Scheerens & Creemers, 1989; Stoll & Fink, 1996; Stringfield et al., 2008a, 2008b; Thoonen, Sleegers, Oort, & Peetsma, 2012). Also implicit in this perspective is the view that the modal practices which are embedded in and reflected by school's culture represent a key mediator of improvement efforts (e.g., Barth, 1990; Fullan, 2001, 2007; Hallinger & Heck, 2011; Hargreaves, 1995; Hargreaves & Fullan, 1998; Potter et al., 2002; Reynolds, 2010; Saphier & King, 1985; Sarason, 1982; Sleegers et al., 2002; Thoonen et al., 2012). Yet, as suggested above, scholars have also noted the importance of subcultures that exist within the larger culture of a given school (Louis & Miles, 1990; McLaughlin, & Talbert, 2001, 2006; Rosenholtz, 1989; Sammons et al., 1997; Schaffer et al., 2012; Siskin, 1991, 1997; Stringfield et al., 2008a).

With these conceptual perspectives in mind, we observe that approaching the improvement of learning for students has a long foundation in policy, research, and practice (Firestone & Corbett, 1988; Hattie, 2009; Hawley & Rosenholtz, 1984). These beliefs found their fullest expression during the 1990s in the design of a wide range of "whole school improvement programs". Unlike incremental school improvement planning approaches, these whole school improvement programs began with the idea of redesigning the school around a coherent set of mutually reinforcing beliefs and practices (Levin, 1987). Examples of these whole school improvement programs included *High Reliability Schools* (Stringfield et al., 2008a, 2008b), *Success for All* (Slavin & Madden, 2006), and *Accelerated Schools* (Lee, Levin, & Soler, 2005; Levin, 1987).

Concurrent with the evolution of these school improvement efforts that target the whole school, however, an alternate literature emerged that focused not on the commonality of the whole but on the uniqueness or variability of its sub-units. This literature can be traced back to Susan Rosenholtz's (1989) seminal paper analyzing the "subcultures" that can be identified in different subject departments located in the same secondary

school. Thus, staff in different departments could demonstrate quite different attitudes towards the workplace, innovation, and change (Rosenholtz, 1989). Rosenholtz and others proposed that in secondary schools the subculture of subject departments could represent a more influential "context" for teachers than the culture of the school as a whole (e.g., Louis & Miles, 1990; Siskin, 1991).

Subsequent efforts have sought to extend this perspective of building secondary school improvement upon departmental change both in the UK (e.g., Hargreaves & Hopkins, 1994; Hargreaves & MacMillan, 1995; Harris, 1998; Harris, Jamieson, & Russ, 1995; Reynolds, 2010; Schaffer et al., 2012; Stringfield et al., 2008a, 2008b) and the USA (e.g., McLaughlin & Talbert, 2001, 2006; Siskin, 1991, 1997). Research conducted within this theoretical perspective has sought to analyze the relationship between secondary school department cultures, teacher workplace behavior, and the impact on school-wide improvement.

For example, Sammons and colleagues (1997) found that English Subject Department Heads' perceptions of staff knowledge of subject content and the General Certificate of Secondary Education (GCSE) syllabus, experience in teaching the subject, and qualifications were significantly correlated with their effectiveness in promoting overall GCSE performance and English performance. They also observed that high levels of teacher absenteeism and turnover, especially in inner-city schools, have knock-on effects on teaching quality and departmental effectiveness. Academically effective secondary schools appeared to benefit from greater coherence between school and departmental cultures. The researchers observed that in these schools the school culture and departmental subcultures appeared to be mutually reinforcing and related to GCSE results.

Reynolds and colleagues applied the concept of "high reliability schooling" (HRS) to the analysis of school improvement processes in secondary schools (Stringfield et al., 2008a, 2008b, 2012). In this context, high reliability schooling refers to strategic efforts aimed at developing and increasing the level of consistency in key education processes as well as the performance outcomes of schools (see Potter et al., 2002; Reynolds, 2010; Schaffer et al., 2012). This program of conceptual and empirical research represents a particularly relevant body of work for scholars and practitioners engaged in understanding the challenges of improving performance in secondary schools (Reynolds, Hopkins, Potter, & Chapman, n.d.; Schaffer et al., 2012; Stringfield et al., 2008a, 2008b). In the course of implementing the "high reliability project" strategies with schools, Reynolds and colleagues (2006) noted, for example:

> [Providing a knowledge base on departmental effectiveness] proved to be particularly powerful, since the department was comprised of people teaching the same subject and therefore had teachers with something in common, and since the department's performance was directly shown in public performance indicator materials that all schools possessed on performance on different subjects. The department was also a manageable entity that was, in practice, an alterable as well as a 'proximal' variable. (p. 64)

Findings from the HRS research support the efficacy of the following strategic elements for improving teaching and learning in secondary schools:

- *clarity of mission*: a small number of clear, agreed, and inflexible goals, with ambitious targets for pupils' academic achievement at their heart;
- *careful monitoring of key systems* to avoid cascading error;
- *data richness*, with good benchmarking and openness about performance data;

- *standard operating procedures* (SOPs), including an agreed model of teaching and consistent implementation of agreed actions in teaching, managing learning behaviour, attendance, etc.;
- *focus on pupils at risk* of failure;
- *pro-active, extensive recruitment and targeted training*, including the delivery of the agreed teaching methods;
- *rigorous performance evaluation* to ascertain the rapid, early, and continuous impact of initiatives;
- *maintenance of equipment* in the highest working order. (Reynolds et al., n.d., p. 6; see also Potter et al., 2002)

Thus, 25 years after Rosenholtz's initial foray into reconceptualizing secondary school improvement, this evolving literature offers empirical support for her proposition that the subject department represents a key focal unit in secondary school improvement (e.g., Harris, 1998; McLaughlin & Talbert, 2006; Sammons et al., 1997; Schaffer et al., 2012; Stringfield et al., 2008a, 2008b). These theoretical perspectives guided the framing of our study of secondary school improvement in Hong Kong. For example, the study incorporated contextual conditions as a set of moderating variables within a longitudinal research design capable of capturing change in department and overall school performance over time (Hallinger & Heck, 2011). From a practical perspective, longitudinal data are needed if we are to gain a better understanding of the patterns of change that occur both within and across schools during the "journey of school improvement" (Hallinger & Heck, 2011; Jackson, 2000; Mulford & Silins, 2009). The paucity of longitudinal data with which to conduct research on improvement across large numbers of schools has, however, until recently stalled progress in identifying and understanding patterns of change in the improvement of schools.

Method

In order to explore secondary school improvement from the dual perspectives of the whole school and subject departments, the current study analyzed student achievement data collected from 47 Hong Kong secondary schools over a 3-year period of time. The diverse set of student achievement data organized by subject departments offered an unusual opportunity to gain insight into patterns of change across departments within these secondary schools. We would also note that our review of the literature found no studies of secondary school improvement from this perspective. Thus, this empirical investigation in Hong Kong offered a unique opportunity to explore these issues outside of the Anglo-American mainstream of school improvement research.

School sample

Forty-seven principals offered access to their school achievement data after invitations to participate in this study were sent to principals of Hong Kong's 498 secondary schools in 2009. The background of the participating schools indicated that they were spread over diverse geographic areas and governed by a broad range of school sponsoring bodies.[1] Although we acknowledge the limitations of this convenience sample, the schools were roughly representative of Hong Kong secondary schools with respect to socioeconomic status, geography, and school board membership.

Given this sample was the first in the literature that provided value-added results in Hong Kong, it is important to check its representativeness based on the distribution of stanines as presented in Table 1. As shown, the distribution of the stanines of sample schools matched closely with the expected frequency, except that there was a bias concerning schools with Stanine 1 results. Between Core 3 stanines and Best 6 subjects stanines, a close match with the latter was desirable because it represented a broader range of subjects than Core 3.

Student achievement data

The data collected for this study consisted of student achievement data collected annually from Hong Kong secondary school students. The achievement data were in the form of value-added scores of students taken from the Hong Kong Certificate of Education Examination (HKCEE), an annually administered standardized examination annually.[2] This examination covers a wide range of key learning areas that correspond to the subjects studied by Hong Kong's secondary school students.

The HKCEE was taken by most Hong Kong students upon completion of their 5-year secondary school education. The exam results, therefore, reflect value-added effects of the 5 years of secondary teaching and learning. While value-added information is often publicly accessible information in the UK or USA, in Hong Kong this is treated as highly confidential information due to inter-school competition for students. Thus, with the schools' permission, we obtained the information directly from the Hong Kong Education Department.

Data analysis

Data extracted from the examination results included two types of metrics: value-added results and group stanine scores. Given their uniqueness and importance to understanding the results of the study, we describe them in detail.

Value-added results

The scores employed in our subsequent data analysis consisted of value-added stanines by subject, rather than the actual value-added estimates. These stanines, ranging from 1 to 9, were scaled from schools' raw value-added estimates. Each value-added stanine corresponds to one of nine intervals of a normal distribution. The intervals represent a width of 0.5 standard deviations excluding the first and the last. In general, Stanines 1 to 3 are categorized as negative value added, Stanines 4 to 6 as no value added, and Stanines 7 to 9 as positive value added. A change in two or more stanines (i.e., representing a change of at least one standard deviation) in any two given year period is generally considered a "significant magnitude of change" rather than a minor fluctuation.[3] Since the introduction of the value-added stanine subject profiles, principals in Hong Kong's secondary schools have used this information to monitor their school-wide learning performance.

Stanine scores for individual subjects and multiple subjects

Between 2006 and 2008, value-added results were potentially available for 21 subjects in six Key Learning Areas (KLAs) in the HKCEE.[4] Although the majority of the schools in our sample offered 16 to 19 subjects, others offered as few as 7 (see Table 2). Beyond the

Table 1. Distribution of value-added stanines of schools.

		Stanine 1	Stanine 2	Stanine 3	Stanine 4	Stanine 5	Stanine 6	Stanine 7	Stanine 8	Stanine 9	Total
Year		*N* (%)	*N* (%)	*N* (%)	*N* (%)	*N* (%)	*N* (%)	*N* (%)	*N* (%)	*N* (%)	*N* (%)
2006 Core 3 subjects	Count (%) by each stanine	***0 (0)*** ↓	***5 (11)*** ↑	***4 (9)*** ↓	***10 (21)***↑	***7 (15)***↓	***10 (21)***↑	6 (13)	2 (4) ↓	3 (6) ↑	47 (100)
2006 Best 6 subjects	Count (%) by each stanine	***0 (0)*** ↓	2 (4) ↓	***9 (20)*** ↑	7 (15) ↓	9 (20)	8 (17)	6 (13)	3 (6)	3 (6) ↑	47 (100)
Expected frequency	Count (%) by each stanine	2 (4)	3 (6)	6 (13)	8 (17)	9 (20)	8 (17)	6 (13)	3 (6)	2 (4)	47 (100)

Notes: Core 3: value-added results based on Chinese, English, and Mathematics; Best 6: value-added results based on any best 6 subjects. Expected frequency represents what the percentage and count would be if the sample was exactly normally distributed. Text in bold and italics indicates larger differences between the frequencies in the sample and expected frequencies. ↑ indicates a frequency higher than expected, while ↓ indicates a frequency lower than expected.

Table 2. The total number of subjects taken in a school and its number of subjects with stanines indicating no value added, positive value added, or negative value added.

Schl Number	1	3	4	5	6	7	8	9	10	11	12	14	15	16	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	Tot
2006 Neg	0	0	6	6	0	1	7	0	**14**	2	7	3	1	**9**	**7**	**11**	2	3	**9**	1	0	0	0	3	7	1	**9**	3	2	2	6	**8**	**12**	4	0	6	**8**	5	**10**	6	5	0	10	0	1	0	0	**11**
2006 Stable	**13**	2	**9**	**13**	**7**	4	**8**	**9**	3	**8**	**8**	**8**	**8**	7	5	1	4	**8**	5	**8**	**15**	7	**11**	**9**	**10**	**10**	4	7	**5**	6	**7**	3	3	**9**	3	**11**	6	**9**	5	**10**	**10**	1	5	**13**	**9**	5	**12**	**27**
2006 Pos	3	**14**	1	0	3	**13**	2	8	0	6	1	5	0	1	4	3	**9**	6	2	7	1	**10**	4	3	1	7	3	**11**	0	**8**	1	0	3	5	**15**	2	3	4	0	1	2	**13**	1	3	**9**	**11**	4	**8**
Total	**16**	**16**	**16**	**19**	**10**	**18**	**17**	**17**	**17**	**16**	**16**	**16**	**9**	**17**	**16**	**15**	**15**	**17**	**16**	**16**	**16**	**17**	**15**	**15**	**18**	**18**	**16**	**21**	**7**	**16**	**14**	**11**	**18**	**18**	**18**	**19**	**17**	**18**	**15**	**17**	**17**	**14**	**16**	**16**	**19**	**16**	**16**	
2007 Neg	1	0	3	6	2	0	2	0	**12**	0	7	2	2	**10**	**8**	**10**	0	4	7	0	0	0	0	3	**12**	2	5	4	**5**	2	**8**	8	**11**	1	0	**9**	**8**	2	**7**	**10**	4	0	**7**	3	1	0	1	**13**
2007 Stable	**13**	4	**8**	**10**	**8**	2	**8**	4	4	6	**8**	**7**	**7**	5	4	4	3	**8**	**9**	5	**15**	**9**	**9**	**12**	4	7	5	6	1	6	7	**10**	6	7	3	8	**8**	**13**	**7**	4	**9**	4	6	**11**	7	**8**	**9**	**24**
2007 Pos	3	**11**	3	1	2	**16**	7	**14**	1	**10**	1	**7**	0	2	3	1	**11**	5	0	**10**	2	7	7	0	1	**8**	5	**9**	1	**8**	0	0	1	**10**	**13**	1	1	2	2	3	3	**12**	3	3	**10**	7	6	**12**
Total	**17**	**15**	**14**	**17**	**12**	**18**	**17**	**18**	**17**	**16**	**16**	**16**	**9**	**17**	**15**	**15**	**14**	**17**	**16**	**15**	**17**	**16**	**16**	**15**	**17**	**17**	**15**	**19**	**7**	**16**	**15**	**18**	**18**	**18**	**16**	**18**	**17**	**17**	**16**	**17**	**16**	**16**	**16**	**17**	**18**	**15**	**16**	
2008 Neg	0	0	1	5	1	0	1	0	6	3	6	4	2	**11**	**8**	**12**	0	1	**9**	1	3	0	2	3	**10**	1	4	1	5	0	5	**8**	**8**	1	0	**12**	4	3	3	2	**8**	0	**8**	8	4	0	0	**10**
2008 Stable	7	3	**13**	**11**	**7**	1	8	7	**8**	**11**	**8**	**9**	**5**	4	5	3	0	4	6	**10**	**12**	3	**10**	6	8	**11**	**9**	8	**8**	7	**9**	**8**	6	1	2	4	6	**12**	**12**	7	7	2	**8**	**9**	**7**	1	**10**	**22**
2008 Pos	**10**	**13**	1	1	4	**17**	**9**	**10**	3	2	2	3	1	0	2	1	**15**	**12**	1	3	1	**13**	3	**7**	0	5	2	**10**	3	**9**	1	0	3	**15**	**14**	2	**7**	2	1	**8**	1	**13**	0	1	**7**	**14**	6	**18**
Total	**17**	**16**	**15**	**17**	**12**	**18**	**18**	**17**	**17**	**16**	**16**	**16**	**8**	**15**	**15**	**16**	**15**	**17**	**16**	**14**	**16**	**16**	**15**	**16**	**18**	**17**	**15**	**19**	**16**	**16**	**15**	**16**	**17**	**17**	**16**	**18**	**17**	**17**	**16**	**17**	**16**	**15**	**16**	**18**	**18**	**15**	**16**	

Notes: Frequency = the frequency of schools with their majority of subjects with stanines indicating stable value-added (Stanines 4– 6), positive value-added (Stanines 7–9), or negative value-added (Stanines 1–3). The School Number ran to 51 because there were missing data for Schools 2, 13, 17, and 35. These were schools which joined in other parts of data collection (teacher survey and student survey) but did not agree to release their value-added data.

Core 3 subjects (i.e., Math, English, Chinese), students typically take between 3 and 7 elective subjects. This depends on their interests, ability, and the availability of subjects in their schools. Given this wide range of subjects offered across schools, assessing the schools by improvement in only a few selected subjects would offer an incomplete picture of their performance.

In addition to value-added stanine results for individual subjects, Hong Kong's testing authorities also provide two composite indicators of achievement results in clusters of subject areas. These composite metrics are referred to as "Core 3 Subjects" (*Core 3 Stanine*) and "Best 6 Subjects" (*Best 6 Stanine*). These measures were employed in this study in order to gain insight into patterns of growth in student achievement across different departments within schools.

The *Core 3 Stanine* metric provides a composite measure of student achievement in three core subjects: Chinese Language, Mathematics, English Language. We refer to these as "Core Subjects" because students are generally required to "pass" these subjects to attain admission to local universities. Given that secondary schools are divided into two streams dependent on their medium of instruction (Chinese or English), the *Core 3 Stanine* metric tends to create an advantage for schools whose medium of instruction (MOI) is English.

The *Best 6 Stanine* represents the value-added stanine for student achievement in the six subjects in which the school's students performed best out of the full set of subjects in which they were tested. Thus, as we look across schools, the *Best 6 Stanine* is actually measuring growth in a different set of subjects for different schools. Despite the lack of comparability in terms of specific subject area achievement, the indicator provides a useful perspective on growth in those subjects in which students are doing best. The Hong Kong Examinations and Assessment Authority (HKEAA) has reported this measure for many years, and over time it has gained acceptance as the most widely used local indicator of *overall performance of secondary schools* by educators and the public.

The *Best 6 Stanine* which measures growth in varied sets of subjects is considered fairer than the *Core 3 Stanine* which is based on a fixed set of subjects. Nonetheless, the *Core 3 Stanine* is an important indicator of school performance in highly regarded subjects. Due to patterns of teacher assignment in Hong Kong during the secondary education, the *Best 6 Stanine* is subject to greater fluctuation than the *Core 3 Stanine*.

Multiple cohort comparisons

As noted earlier, longitudinal data offer useful advantages when examining school improvement. However, longitudinal data can be structured in different ways. Willms (1992) suggested that, "growth measures are more reliable if they are based on measurements taken on at least three occasions" (p. 36).

In the present study, value-added stanines were based on estimates of the final exam of Primary 6 and the HKCEE exam separated by 3 years. We assessed school improvement by comparing the value-added stanines of the cohort in 2006 with those of other cohorts. For example, we employed the value-added *Best 6 Stanine* of a school in 2006 as an indicator of the overall performance of the school's cohort in 2006. Then, we also compared it with the value-added stanines of 2007 and 2008 to assess the performance of different cohorts. Thus, for a school with a stanine of 6 in 2006 and a stanine of 7 in 2008, this small but positive difference in stanine can be regarded as an indication of school improvement of cohorts between 2006 and 2008. This approach contrasts with

Hallinger and Heck's (2011) longitudinal study of school improvement where change in student learning outcomes was estimated for a *single cohort* over a 4-year period.

A limitation of stanines revolves around their tendency to fluctuate from year to year. We therefore considered 3 years as a minimum period over which to monitor growth in the performance of cohorts. Although this observation window of 3 years still does not eliminate potential bias caused by annual fluctuation, it has compensating advantages when compared with evaluating school improvement based on a single cohort.

Analytical procedures

Analysis of these achievement data began with producing a set of descriptive tables showing year-by-year patterns of growth, stagnation or coasting, and decline. These descriptive data were employed to gain insight into relative growth of the schools versus their composite departments, as well as magnitude of change over time for different departments. Subsequently, Pearson's test was used to examine patterns of change for Core 3 subjects. Multiple regression was used to examine how different patterns of subject department change could account for whole school improvement over time.

Results

As stated earlier, the first objective of this study was to determine whether identifiable patterns can be established in the improvement of subject departments over time. Here, we were interested both in patterns of improvement, stagnation or decline, as well as the magnitudes of changes in department performance over time. Following these analyses, we examine how changes in subject department performance compare to improvements in overall school performance in our sample of Hong Kong secondary schools.

Analysis of school versus department performance

To compare changes in whole school and subject department performance over time, we began our analysis by examining patterns of change among subject departments in the 47 schools at three points in time (i.e., 2006, 2007, 2008). The results, shown in Table 2, yielded a wide variety of patterns of change in the performance of subject departments within individual schools over time. In order to make sense of these patterns, we can group schools based on two characteristics:

(1) *values* (i.e., *positive, neutral*, or *negative*) of the value-added stanines of the majority of subjects in a school; and
(2) *frequency* of change in the direction of value-added stanines over the 3 years.

When we group patterns of change according to these two criteria, a variety of different trajectories in subject department growth emerged:

- schools that showed a trajectory with positive value-added stanines consistently in most subjects over 3 years (i.e., Schools 3, 7, 20, 31, 33, 39, 46);
- schools that showed a trajectory with positive value-added stanines gradually in most subjects in 2008 (i.e., Schools 1, 8, 9, 21, 25, 27, 41, 49, 50);
- schools that showed a trajectory with neutral value-added stanines consistently in most subjects over 3 years (i.e., Schools 4, 5, 6, 12, 14, 15, 24, 26, 42, 48, 51);

- schools that showed a trajectory with neutral value-added stanines gradually in most subjects in 2008 (i.e., Schools 10, 11, 23, 29, 30, 32, 34, 43);
- schools that showed a trajectory with negative value-added stanines consistently in most subjects over 3 years (i.e., Schools 16, 18, 19, 37, 47);
- schools that showed a trajectory with negative value-added stanines gradually in most subjects in 2008 (i.e., Schools 22, 28, 36, 40, 45);
- school that showed an inconsistent trajectory with value-added stanines in most subjects over 3 years (i.e., School 44).

Magnitude of change in individual subjects

The magnitude and scope of whole school improvement in secondary schools is dependent upon the magnitude of improvement of individual subjects and the scope of subjects studied. That is, variability in the performance of departments gives shape to the density or coherence of school-wide improvement. For example, the inclusion of technology subjects such as Computer and IT, Commerce, and Word Processing & Business Communication (English) are often expected to boost overall school performance. Local educators often assume that achieving good results in these subjects is "easier" than in academic subjects like English or Physics.

Because the magnitude and scope of improvement may vary across individual subjects, the selection of "subjects" is critical for understanding their contribution to school-wide improvement. To facilitate further comparisons, we selected 11 subjects that were widely available in the schools comprising our sample. We summarized the change in stanine scores among these 11 subjects between 2006 and 2008 in terms of two composite measures that we refer to as *Best 6 Stanine* and *Core 3 Stanine* (see Tables 3 and 4).

The results in Table 4 confirm that variability in departmental performance varies sharply across different subjects. For most schools and for all subjects and composite measures, changes in stanines tended to fall within the +/− 2 range. As noted earlier, we consider a change of stanine of two or more stanines as representing a "significant" magnitude of change.

The magnitude of gain or loss in a subject in a particular school can be quite substantial. For example, we noted improvements as large as seven stanines (e.g., in History) and declines as large as five stanines (e.g., Chinese). Significant declines (e.g., greater than one stanine; Column D in Table 4) were, however, limited to Chinese, Physics, and Chinese History. Moreover, these significant declines occurred in less than 15% of the schools in our sample.

In contrast, significant *positive* changes were found in more than 15% of the sample schools and across almost all subjects (see Column E in Table 4). Substantial improvements were most frequently found in Chinese, English, Chemistry, and Accounting, with significant positive changes reaching 29%. What is interesting is that, although English is a foreign language, it is relatively easier to obtain a large positive change of two or more stanines in English than in Chinese across cohorts. The data did not support the common local belief that improvements are more easily achieved in Math than in English. Indeed, significant positive gains in Math were seldom achieved by different cohorts within the schools in our sample.

As we noted earlier, comparing the results of different cohorts offers insight into the magnitude of change over time. Column C in Table 4 shows the results indicating the difference in the percentage of schools with a small positive change or small negative change. It is interesting that the schools that showed a small positive gain between the

Table 3. Changes in stanine scores in 11 subjects between 2006 and 2008.

Subject	Count	% of schools with stanine change of 7	% of schools with stanine change of 6	% of schools with stanine change of 5	% of schools with stanine change of 4	% of schools with stanine change of 3	% of schools with stanine change of 2	% of schools with stanine change of 1	% of schools with stanine change of 0	% of schools with stanine change of –1	% of schools with stanine change of –2	% of schools with stanine change of –3	% of schools with stanine change of –4	% of schools with stanine change of –5
Chinese	46	0%	0%	2%	0%	7%	20%	20%	17%	13%	15%	4%	0%	2%
English	45	0%	0%	0%	0%	2%	20%	36%	18%	18%	4%	2%	0%	0%
Math	46	0%	0%	0%	0%	2%	11%	30%	28%	17%	11%	0%	0%	0%
Physics	43	0%	0%	0%	0%	9%	12%	23%	21%	19%	16%	0%	0%	0%
Chemistry	44	0%	0%	0%	2%	9%	14%	23%	23%	25%	5%	0%	0%	0%
Biology	44	0%	0%	5%	0%	2%	14%	32%	27%	11%	7%	2%	0%	0%
Economics	43	0%	0%	0%	2%	7%	9%	16%	30%	26%	5%	5%	0%	0%
Geography	44	0%	0%	0%	0%	2%	16%	27%	23%	18%	7%	7%	0%	0%
History	36	3%	0%	0%	0%	0%	14%	19%	22%	31%	11%	0%	0%	0%
Chi. History	43	0%	0%	0%	5%	0%	14%	28%	16%	12%	16%	5%	5%	0%
Accounting	31	0%	0%	0%	0%	16%	13%	10%	23%	26%	13%	0%	0%	0%
Core 3	47	0%	0%	0%	0%	0%	15%	34%	28%	13%	11%	0%	0%	0%
Best 6	47	0%	0%	0%	0%	4%	15%	26%	30%	19%	6%	0%	0%	0%

Table 4. Overall positive and negative changes in stanine scores in 11 subjects and 2 composite measures between 2006 and 2008.

		A	B	C	D	E	F
Subject	*N* of schools offering subject	% of schools with negative change	% of schools with positive change	Diff. in % of schools with + and – change	% of schools with ≤ –2 change	% of schools with ≥ +2 change	Diff. in % of schools with ≥ +2 and ≤ –2 change
Chinese	46	35%	48%	13%	22%	28%	7%
English	45	24%	58%	33%	7%	22%	16%
Math	46	28%	44%	15%	11%	13%	2%
Physics	43	35%	44%	9%	16%	21%	5%
Chemistry	44	30%	48%	18%	5%	25%	21%
Biology	44	21%	52%	32%	9%	21%	11%
Economics	43	35%	35%	0%	9%	19%	9%
Geography	44	32%	46%	14%	14%	18%	5%
History	36	42%	36%	–6%	11%	17%	6%
Chi. History	43	37%	47%	9%	26%	19%	–7%
Accounting	31	39%	39%	0%	13%	29%	16%
Core 3	47	23%	49%	26%	11%	15%	4%
Best 6	47	26%	45%	19%	6%	19%	13%

2006 and 2008 cohorts exceeded those that showed a negative decline by about one third in English and Biology. We also compared the difference in the percentage of schools with a significant positive change and the percentage of schools with a significant negative change (see Column F in Table 4). History was the only subject that schools would find it easier to decline than improve. In contrast, the sample schools were more likely to improve than decline significantly in Chemistry and English between cohorts. When comparing the results of Column C and Column F, we also found the likelihood to have a significant change rather than a minor change was higher in Chemistry (21% vs. 18%) and Accounting (16.1% vs. 0.0%).

Magnitude of change on composite measures

In general, the variability of composite measures of school-wide improvement was smaller than those for individual subjects. As shown in Column C, almost 20% of the sample schools had a significant positive change of more than one stanine between 2006 and 2008 for *Best 6*, but only about 15% for *Core 3*. Results in Column B show that it is more likely for schools in the sample to have a significant negative change in *Core 3* (10.6%) than in *Best 6* (6.4%). Thus, the overall likelihood or the net difference to have a significant positive change rather than a significant negative change (i.e., Column D) for *Best 6* (12.8%) is about 3 times of that for *Core 3* (4.3%).

An interesting finding is that when small changes including one stanine are taken into account (Column A), there were about 26% more schools that would have a positive stanine change than a negative stanine change for *Core 3*, but only about 19% for *Best 6*. These results suggest that it would be more difficult to achieve a significant positive gain in *Core 3* achievement than in *Best 6*.

Variability of improvement in individual schools

The results shown in Tables 2 and 3 affirmed that different subject departments within a school can have positive, negative, or neutral stanines for the same cohort. The discrepancies between Column C and Column F in Table 4 indicate that the net difference in changes in stanine for subjects can be different when minor changes or significant changes are considered. These results suggest two refinements in our conceptualizations of improvement in secondary schools: (a) scope of school improvement and (b) magnitude of school improvement.

We can conceptualize the *scope of school improvement* in terms of the net difference in the percentage of subjects with a positive change and subjects with a negative change across two cohorts (see the fifth row in Table 5). This perspective suggests that in order to achieve school-wide school improvement, it may not be sufficient to have positive stanine changes in only a few departments. Rather, it is necessary to create a critical mass of departments with improving stanine scores over time (e.g., see Schools 7, 38). When improvements are localized in only a few departments, "school improvement" is unlikely to be "felt" by the stakeholders in the school (e.g., see Schools 3, 11, and 24).

Second, in concert with considering the scope of improving departments, we can also examine the *magnitude* of growth in subject area achievement across the school. Here, magnitude can be seen in the net difference in percentage between subjects with significant positive changes and subjects with significant negative changes, as indicated in the last row in Table 5. Again, the former has to outnumber the latter. School 7 is a typical case where the net percentage of overall changes of all subjects was large (80%), but its net percentage of significant changes in all subjects was small (10%). School 31 is a typical contrasting example.

Now we have four indicators of school improvement at the school level varying in terms of the scope and magnitude of change, but fixed between 2006 and 2008: *Net Overall Change, Net Significant Change*, *Change in Best 6 Stanine*, and *Change in Core 3 Stanine*. The relationships of these indicators are plotted in Figure 1.

The bubble chart in Figure 1 shows a linear function between the *Net Overall Change* and the *Net Significant Change*, suggesting a strong correlation between them. The bubbles rise from the left bottom quadrant to the right top quadrant forming a clear straight line. This suggests not only that a significant positive change in stanine contributes to the overall school improvement but also that the more subjects with this kind of changes the better. The relationship between the *Net Overall Change* and the *Best 6 Stanine Change* is also evident in that no bubbles with colored shades indicating a negative change appear in the quadrants on the right, while only one bubble (School 40) indicating a small *Best 6 Stanine Change* also has a small negative *Net Overall Change*.

This slightly weaker relationship is also indicated by that schools with the largest positive *Best 6 Stanine changes* (the biggest bubble; Schools 8 and 44) do not have the largest *Net Overall Change*. This result suggests that the *Net Overall Change* is a broader and thus a better indicator of school-wide improvement as it covers 11 subjects while the *Best 6 Stanine Change* reflects information on only 6 subjects. A weaker relationship between the *Net Overall Change* and the *Core 3 Stanine Change* is also noted. Several schools (i.e., Schools 14, 16, 19, 26, and 40) had a positive *Core 3 Stanine Change*, but their *Net Overall Changes* were all negative. School 33 is the only school with a negative *Core 3 Stanine Change* but a positive *Net Overall Change*. These results suggest that while it is less common to have a negative *Core 3 Stanine Change* and a positive *Net*

Table 5. Variability in the magnitude of subject area improvement in individual secondary schools.

	School No./ Change	1	3	4	5	6	7	8	9	10	11	12	14	15	16	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51
1	**% subjects with neg change**	27	50	40	55	14	0	9	22	0	73	22	30	67	44	40	40	20	0	40	27	91	27	60	0	55	46	10	36	0	18	36	14	46	0	50	46	20	30	0	9	46	13	36	60	46	22	50
2	**% subjects with pos change**	46	40	40	0	43	80	73	67	91	18	44	30	33	22	40	40	60	91	40	64	9	46	20	50	9	9	60	64	50	55	27	57	46	100	10	55	60	50	73	73	18	75	46	0	27	44	40
3	**% net changes of all subjects**	18	**10**	0	55	29	80	64	44	91	**55**	22	0	**33**	**22**	0	0	40	91	0	36	**82**	18	**40**	50	**46**	**36**	50	27	50	36	−9	43	0	100	**40**	9	40	20	73	64	**27**	63	9	**60**	18	22	**10**
4	**% subjects with ≤−2 change**	9	10	0	27	0	0	0	11	0	55	11	10	0	33	30	30	0	0	40	9	36	0	30	0	27	18	10	0	0	0	9	0	18	0	40	9	0	0	0	0	18	13	0	30	46	0	0
5	**% subjects with ≥+2 change**	18	20	10	0	14	10	55	22	55	0	33	10	0	22	20	10	40	64	20	46	0	18	10	0	9	0	20	36	0	18	9	43	18	55	0	27	30	30	36	46	0	13	0	0	18	33	10
6	**% net sig change in all subjects**	9	10	10	**27**	14	10	55	11	55	**55**	22	0	0	**11**	**10**	**20**	40	64	**20**	36	**36**	18	**20**	0	**18**	**18**	10	36	0	18	0	43	0	55	**40**	18	30	30	36	46	**18**	0	0	**30**	**27**	33	10

Note: All values are percentages. Values highlighted in bold font are negative.

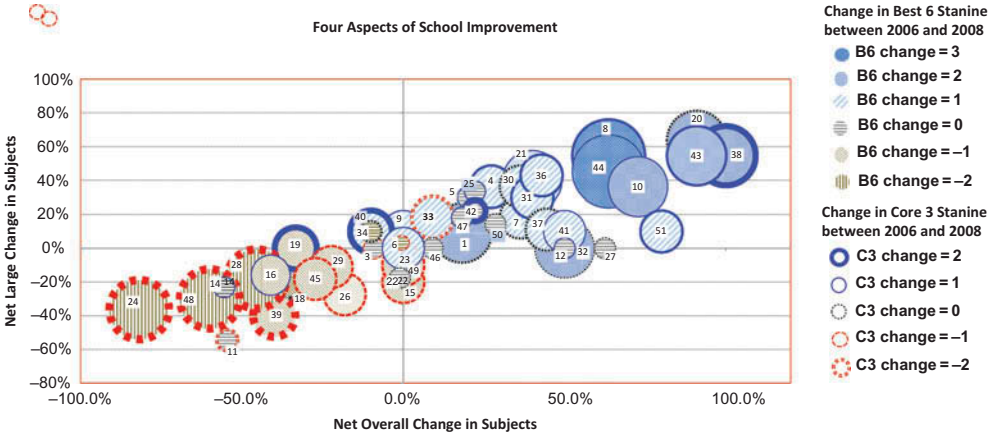

Figure 1. Four perspectives on school improvement: Net Overall Change, Net Large Change, Change in Best 6 Stanine, and Change in Core 3 Stanine.

Overall Change concurrently, a positive *Core 3 Stanine change* does not necessarily predict a positive *Net Overall Change*.

This implies that improvement in *Core 3* subjects is a necessary but insufficient condition for producing overall school improvement. Finally, we noted only a weak positive relationship between the *Best 6 Stanine Change* and the *Core 3 Stanine Change* because there are many exceptions where negative *Best 6 Stanine Changes* are connected with positive *Core 3 Stanine Changes* or vice versa (e.g., Schools 16, 19, 25, 26, and 33). There are also exceptions in which positive *Best 6 Stanine Changes* involve no *Core 3 Stanine Changes*, suggesting many schools can manage positive stanine changes in subjects other than the *Core 3* subjects (e.g., Schools 1, 7. 20, 30, 32, 37).

A Pearson correlation test was performed to explore the relationships among these indicators. The results, summarized in Table 6, were compatible with the patterns identified in Figure 1. There was a strong correlation between the *Net Overall Change* and the *Net Significant Change* with r = .833, significant at <0.00 level (1-tailed). The correlation between the *Net Overall Change* and the *Best 6 Stanine*

Table 6. Correlations among four indicators of school improvement.

Perspectives on School Improvement	Net Overall Change in Subjects	Net Large Change in Subjects	Core 3 Stanine Change 2006–08	Best 6 Stanine Change 2006–08
Net Overall Change in Subjects	1.000			
Net Large Change in Subjects	.833	1.000		
Core 3 Stanine Change between 2006 and 2008	.499	.569	1.000	
Best 6 Stanine Change between 2006 and 2008	.798	.777	.461	1.000

Note: N = 47, all correlations significant at ≤0.001 level (1-tailed).

Table 7. Multiple regression results of an overall school improvement model.

Model DV: Net Overall Change $N = 47$	Unstandardized coefficients		Standardized coefficients	
	B	Std. error	Beta	Sig.
(Constant)	.022	.035		.519
Net Significant Changes in Subjects	.823	.205	.523	.000**
Core 3 Stanine Change between 2006 and 2008	.010	.034	.026	.779
Best 6 Stanine Change between 2006 and 2008	.132	.042	.380	.003*

Change was also strong with $r = .798$. A moderately strong relationship with $r = .499$ was found between the *Net Overall Change* and the *Core 3 Stanine Change*, while the weakest correlation was found between the *Best 6 Stanine Change* and the *Core 3 Stanine Change* with $r = .461$.

To explore how the three indicators, *Net Significant Change*, *Best 6 Stanine Change*, and *Core 3 Stanine Change*, may explain the variability of overall school improvement (i.e., *Net Overall Change),* a multiple regression was performed (see Table 7). The regression model ($F(3,43) = 43.465$, $p < .001$) obtained was not only significant but also had a high adjusted R square = .735. As expected, both the *Net Significant Change* ($\beta = .523$, $t = 4.02$) and the *Best 6 Stanine Change* ($\beta = .38$, $t = 3.15$) were the significant predictors, positively contributing large portions of the variance, 27.3% and 18.7%, respectively. These results provided additional evidence to confirm that the more widespread the large departmental improvements are, the larger the overall school improvement will be. Moreover, the *Best 6 Stanine* and its changes are reliable indicators of overall school improvement, while the *Core 3 Stanine* and its changes are not.

Discussion

This study was undertaken to address a gap in the literature on secondary school improvement. Along with other scholars studying school improvement in secondary schools (e.g., Louis & Miles, 1990; McLaughlin & Talbert, 2001; Rosenholtz, 1989; Sammons et al., 1997; Schaffer et al., 2012; Siskin, 1997; Stringfield et al., 2008a, 2008b), we suggested that the predominant emphasis on "whole school improvement" in the literature could overlook variability of performance among subject departments within secondary schools. Thus, this study was undertaken to provide a detailed analysis of patterns of change in subject department performance in a sample of Hong Kong secondary schools. In this concluding section of the paper, we highlight key findings and implications of this analysis and revisit limitations of the study.

Limitations of the study

Several limitations attended this study of secondary school improvement. First, we note that, unlike typical quantitative studies of school improvement (e.g., Geijsel et al., 1999, 2001; Hallinger & Heck, 2011; Sleegers et al., 2002; Thoonen et al., 2012), we did not examine the contribution of alterable school factors to school improvement. Although this is of practical interest, the lack of attention to patterns of change in the learning outcomes of different departments represented the gap in research and practice of this study. In order

to more clearly illustrate and understand patterns of department performance, we, therefore, limited the scope of the current study to an examination of change in the learning outcomes of subject departments. Subsequent papers will examine the relationship of alterable school-level variables to these performance patterns.

Second, our study focused on a non-random sample of secondary schools in Hong Kong. While the sampling impacts the generalizability of the results, our sample did not seem to be biased. It represented a broad range of school sponsoring bodies and spread over different geographic areas of Hong Kong. It was also representative of the spectrum of socioeconomic contexts (see Walker & Ko, 2011). As noted, our sample was slightly underrepresented by schools with very low value-added scores.

Third, unlike most studies in the school effectiveness and school improvement literature, our value-added data were based on stanines rather than estimates. Stanines are used more often in student assessment especially when comparisons are made between test scores of different tests, across different content areas, and different times. It is considered as a good heuristic measure that indicates meaningful but broad differences in scores. It should, however, be used with caution when fine distinctions in performance are expected (Hills, 1983). The use of stanines suited our purpose to compare the examination results of different cohorts in different subjects for about 10% of Hong Kong secondary schools. Stanines offered more meaningful interpretations than actual value-added estimates when they were not fully available.

The major limitation of stanines is that they may be too crude for some statistical purposes, because two value-added estimates in the same stanine can be further apart than two value-added estimates in adjacent stanines. Although any categorization of school performance based on stanines is a statistically imperfect estimation, stanines and their changes can be useful categories for characterizing schools, for example, as stagnated versus improving schools (Walker & Ko, 2011).

However, the distribution information of stanines is broad because the number of schools in each stanine is not equal but varies with its location in the distribution. For example, the number of schools with a stanine of 1 is about 4%, while that for 5 is 20%. In practice, an improvement from Stanine 8 to Stanine 9 is not comparable in magnitude, nor equally achievable as an improvement from Stanine 1 to Stanine 2, though both involve a difference of 1 in stanine. This suggests that implications of changes in stanine scores could vary depending upon the "initial state" for comparisons. For example, the schools that initially had negative value-added estimates or stanines below 4 would have different expectations and strategies than schools that had already achieved positive value-added estimates or stanines beyond 6.

These findings are generally consistent with those of other scholars who have recently investigated school-level change in learning outcomes over time (e.g., Day et al., 2011; Hallinger & Heck, 2011). Although our study did not examine explanatory factors, these other researchers have suggested that leadership strategies employed by principals in school improvement are influenced by trajectories of change. Despite the limitations of using value-added stanines, they are often used in practice by school systems for monitoring school performance and offer a useful perspective on change in departmental performance (Willms, 1992).

Interpretation of the findings

Research arising from Rosenholtz's (1989) analysis of improvement in secondary schools provided a conceptual perspective and foundation for the current study. Using value-added

data across a variety of subjects in 47 secondary schools, we have provided evidence showing that secondary schools tend to be comprised of subject departments with differential effectiveness and capacity to improve. Our sample schools also tended to evidence different improvement trajectories defined in terms of the value-added stanines in the majority of its subjects within a given period. Neither composite measure, the *Core 3 Stanine* and the *Best 6 Stanine*, could inform individual departmental effectiveness without bias. Conversely, the performance trajectories of individual departments is a poor predictor of school-wide improvement.

Neither changes in *Best 6 Stanine* nor changes in *Core 3 Stanine* provide a complete or accurate picture of whole school improvement. Instead, the extent of variability of school improvement in individual schools is better explored through two indicators, the *Net Overall Change* and the *Net Large Change*. These represent new metrics in the school effectiveness and school improvement research literature. These two indicators showed a significant positive linear relationship with both the *Best 6 Stanine Change* and the *Core 3 Stanine Change*. As expected, the *Net Large Change* is a better indicator of the overall school improvement when it was defined in terms of the *Net Overall Change*.

Trajectories and indicators of school improvement

School improvement is a context-sensitive phenomenon that reflects the unique developmental path of a school (Hallinger & Heck, 2011; Jackson, 2000; Potter et al., 2002; Schaffer et al., 2012). We have considered magnitude, scope, and time periods of change as factors that contribute to school improvements. Our dataset permitted us to compare whole school performance with that of individual subject departments over 3 years, and we were able to identify a variety of trajectories of improvement.

Time is an important factor in school improvement since significant organizational changes do not take place overnight, and often take 3 to 5 years to unfold (Fullan, 2007; Hall & Hord, 2002). Consistency in both processes and results is both an indicator and outcome of successful school improvement (Reynolds, 2010; Stringfield et al., 2008a, 2008b). Few schools (i.e., only 6 out of 47) in our sample showed a trajectory with positive value-added stanines consistently in most subjects over 3 years. Only 10 additional schools managed to have most subjects with a positive value-added stanine in 2008. The majority of schools (40%) were in a trajectory with neutral value-added stanines in most subjects during this time period. These results suggest the difficulty of leading and sustaining school-wide improvement across departments in secondary schools (Louis & Miles, 1990; Schaffer et al., 2012; Stringfield et al., 2008a, 2008b).

Other data in our study offered insight into the time needed to bring about broad-based improvements. The majority of schools in our sample have been improving in most subjects at a pace either as expected (i.e., Stanines 4–6) or somewhat better than expected (i.e., Stanines 7–9). If, however, policymakers are looking for significant improvements within a short time frame, few schools would have met this expectation.

Our results also provide clear evidence that research which relies solely upon student outcomes in math and literacy in secondary schools will provide a distorted and incomplete picture of school effectiveness and improvement. The implicit assumption that improvements in selected subjects will predict broader performance improvements across the "whole school" was not supported in our data. Indeed, our data suggested that the task of achieving improvement in different subjects also varies in difficulty.

Comparing to the *Core 3 Stanine Change* and the *Best 6 Stanine Change,* the two popular indicators of overall school improvement among the local practitioners, the *Net*

Overall Change and the *Net Large Change*, were found better lenses to look at school-wide school improvement. It is important to find that the *Net Large Change* is a better predictor of overall school improvement than the *Best 6 Stanine Change*. This suggests that collective improvements in different subject departments in a school have to show a positive net difference in changes. Between the *Best 6 Stanine* and the *Core 3 Stanine,* one may expect bringing out effective positive changes in the *Core 3* subjects is harder as the departments of *these* subjects are practically larger.

Effective changes in these departments thus naturally involve more teachers in a school. Indeed, the overall likelihood or the net difference to have a significant positive change rather than a significant negative change for the *Best 6 Stanine* (12.8%) was about 3 times of that for the *Core 3 Stanine* (4.3%). However, our results indicated that the *Best 6 Stanine* was a better indicator of overall school improvement. There was only one case (i.e., School 33) in which a school with a positive *Best 6 Stanine Change* also experienced a negative *Core 6 Stanine Change,* but a significant positive *Core 3 Stanine Change* did not warrant a better chance of a positive *Best 6 Stanine Change* (i.e., Schools 14, 16, 19, and 26).

Differential departmental cultures, effectiveness, and change

Our study was not designed so as to directly test Rosenholtz's (1989) thesis. Nonetheless, our data do suggest that subject departments represent an influential context for teachers. Moreover, the outcome data further suggest that department cultures may be more influential than the culture of the school as a whole. Consistent with Sammons et al. (1997), we found that there can be effective departments in ineffective schools, though more effective departments tended to be found in more effective schools. These results also support the emergent findings from studies of high reliability secondary schools (Schaffer et al., 2012; Stringfield et al., 2008a, 2008b).

Comparing the department cultures and effectiveness of different schools is not a straightforward matter. Neither is comparing the culture and effectiveness of different departments in the same school. In general, subject departments of the *Core 3* subjects are large departments consisting of 6 to 12 teachers as they are taught throughout the secondary school years. Subjects like Economics, Accounting, and the Science subjects are small departments of 2 to 4 teachers because they are taught in senior secondary years only. As mentioned, leading changes in *Core 3* subjects can be a more challenging task as they are larger departments. It is interesting that none of these assumptions were supported in the current data.

There were slightly more schools with a positive *Core 3 Stanine* (48.9%) than a *Best 6 Stanine* (44.7%). Subjects that were more likely to have a negative, rather than a positive, stanine change were electives like History, Chinese History, Economics, or Physics. It is not clear whether these results were affected by the fact that our sample schools tended to be schools that experienced school improvements rather than declines. It should be noted that in most Hong Kong secondary schools, only the academically able students are allowed to opt for science subjects as their electives while academically less able students have to take the subjects Principles of Accounting, Economics, or Chinese History.

Between Math and English, the two most studied subjects in school effectiveness and school improvement research, improvements in English were surprisingly more easily achievable than those in Math were over different cohorts in our sample. In Hong Kong, improvement in English enhances the school image of a secondary school and its attractiveness to capable incoming students. For example, School 7 was originally a

Chinese Medium of Instruction (CMI) school with high value-added stanines for 3 consecutive years. It became an English Medium of Instruction (EMI) school in 2010. In interviews, principal and key staff confirmed that the vision to be an EMI school had been at the core of their school improvement strategy. The principal reported that beside mutual trust with teachers, improving the physical environment, setting ambitious targets for better student outcomes, and extending leadership to key staff had been her major strategies for school improvement. These measures have also been identified as contributing to successful school improvement in England (Day et al., 2011; Reynolds, 2010; Schaffer et al., 2012; Stringfield et al., 2008a, 2008b), USA (Bryk et al., 2010; Hallinger & Heck, 2010), as well as Australia (Gurr, Drysdale, & Mulford, 2005; Mulford & Silins, 2009).

In conclusion, this study of change in the performance outcomes of departments and secondary schools in Hong Kong affirms three key directions for future research and practice. First, it affirms the practical utility of approaching school improvement in secondary schools with a more refined and nuanced lens that takes account of their structural and cultural differentiation. Consistent with recommendations of other scholars who have worked these same fields (e.g., Harris, 1998; Louis & Miles, 1990; Sammons et al.,1997; Stringfield et al., 2008a, 2008b), our results argue for leaders to formulate strategies that target departments as a "proximal" target (Reynolds et al., 2006, p. 64) for improvement in secondary schools. Second, it suggests the need for more sophisticated theory development related to the organizational conditions that bear upon successful school improvement in secondary schools. Hearkening back to the seminal work of Rosenholtz (1989) leads us to note the relatively sparse theoretical contributions to this literature in the ensuing decades. Finally, given the delimited scope of this study on performance outcomes, we encourage researchers to undertake studies that examine how school- and classroom-level factors contribute to patterns of change in subject department and school performance. Moreover, consistent with other recent scholarly work in this field (e.g., Hallinger & Heck, 2011; Mulford & Silins, 2009; Stringfield et al., 2008a; Thoonen et al., 2012), we propose that quantitative research on school improvement in secondary schools should adopt longitudinal rather than cross-sectional research designs. Only in this way will the field build a knowledge base that goes beyond description and which is capable of contributing to theory, policy, and practice.

Acknowledgements

The authors wish to thank Professor Ronald H. Heck for his useful comments on an earlier draft of this manuscript.

Funding

The authors wish to acknowledge the funding support of the Research Grant Council (RGC) of Hong Kong for its support through the General Research Fund [GRF 451407].

Notes

1. According to the most recent data available, there are only 457 schools in the major districts after some school closure. Our 47 schools represented about 7%, 8%, 7%, and 10% of these districts. There are 77 fee-paying schools receiving the government direct subsidies, but only 3 were in our sample. All other 36 schools are aided by the government and offering free education to students. Regarding school sponsoring bodies, the 47 schools represented 33

sponsoring bodies with 5 schools from Hong Kong Catholic Diocesan, the largest School Sponsoring Body in Hong Kong. Our sample may slightly underrepresent schools from the Anglican Church and the Chinese Christ Church, but we do not consider this to exercise a substantial impact on our findings.

2. This test is now replaced by the first local standards-referenced reporting of assessments known as Hong Kong Diploma of Secondary Education (HKDSE) Examination in 2012 for students completing 6 years of secondary education under the new academic structure.
3. For example, when an initial value-added stanine is Stanine 1, 4, or 7, a change in stanine of 2 would not affect the categorization of the stanines because the new Stanine 3, 6, or 9 is within the same categorization as the previous one.
4. Examinations were also offered for German, French, and others without value-added results. There were two syllabuses for English Language in 2006 and 2008. Some schools may offer two syllabuses at the same time, so there would be different stanines for the two syllabuses. The six KLAs and their subjects are as follows: (1) Chinese Language Education: Chinese Language, Chinese Literature, and Putonghua; (2) English Language Education: English Language; (3) Mathematics Education: Mathematics and Additional Mathematics; (4) Personal, Social, and Humanities Education: Chinese History, Geography, History, Economics, Integrated Humanities, Religious Studies, and Travel and Tourism; (5) Science Education: Biology, Chemistry, and Physics; and (6) Technology Education: Computer and IT, Commerce, Principles of Accounting, and Word Processing & Business Communication (English).

Notes on contributors

Dr. Philip Hallinger is the Joseph Lau Chair Professor of Leadership and Change in the Hong Kong Institute of Education.

Dr. James Ko is Assistant Professor of Educational Policy and Leadership in the Hong Kong Institute of Education.

Dr. Allan Walker is the Joseph Lau Chair Professor of International Educational Leadership and Dean of the Faculty of Education and Human Development in the Hong Kong Institute of Education.

References

Ainley, J. (1994, January). *Curriculum areas in secondary schools: Differences in student response*. Paper presented at the International Congress for School Effectiveness and Improvement, Melbourne.

Barth, R. (1990). *Improving schools from within*. San Francisco, CA: Jossey-Bass.

Brookover, W., & Lezotte, L. (1979). *Changes in school characteristics coincident with changes in student achievement*. East Lansing: Michigan State University, College of Urban Development.

Bryk, A. S., Sebring, P. B., & Allensworth, E. (2010). *Organizing schools for improvement: Lessons from Chicago*. Chicago, IL: University of Chicago Press.

Crandall, D., Eiseman, J., & Louis, K. S. (1986). Strategic planning issues that bear on the success of school improvement efforts. *Educational Administration Quarterly, 22*(3), 21–53.

Creemers, B. P. M., & Kyriakides, L. (2008). *The dynamic model of educational effectiveness*. New York, NY: Routledge.

Creemers, B. P. M., & Scheerens, J. (1994). Developments in the educational effectiveness research programme. *International Journal of Educational Research, 21*, 125–140.

Cuban, L. (2013). Why so many structural changes in schools and so little reform in teaching practice? *Journal of Educational Administration*. *51*, 109–125.

Day, C., Sammons, P., Hopkins, D., Harris, A., Leithwood, K., Gu, Q., … Kingston, A. (2011). *Successful school leadership: Linking learning and achievement*. Berkshire: Open University Press.

Edmonds, R. (1979). Effective schools for the urban poor. *Educational Leadership, 37*, 15–24.

Edmonds, R. (1982). Programs of school improvement: An overview. *Educational Leadership, 40* (3), 4–11.

Firestone, W. A., & Corbett, H. D. (1988). Planned educational change. In N. Boyan (Ed.), *Handbook of research on educational administration* (pp. 321–340). New York, NY: Longman.
Fullan, M. (2001). *Leading in a culture of change*. San Francisco, CA: Jossey-Bass.
Fullan, M. (2007). *The new meaning of educational change* (4th ed.). New York, NY: Teachers College Press.
Geijsel, F., Sleegers, P., & Van den Berg, R. (1999). Transformational leadership and the implementation of large-scale innovation programs. *Journal of Educational Administration, 37*, 309–328.
Geijsel, F., Sleegers, P., Van den Berg, R., & Kelchtermans, G. (2001). Conditions fostering the implementation of large-scale innovation programs in schools: Teachers' perspectives. *Educational Administration Quarterly, 37*, 130–166.
Gurr, D., Drysdale, L., & Mulford, B. (2005). Successful principal leadership: Australian case studies. *Journal of Educational Administration, 43*, 539–551.
Hall, G. (2013). Evaluating change processes: Assessing extent of implementation (constructs, methods and implications). *Journal of Educational Administration, 51*, 264–289.
Hall, G., & Hord, S. (2002). *Implementing change: Patterns, principles, and potholes*. Boston, MA: Allyn & Bacon.
Hallinger, P., & Heck, R. H. (2010). Collaborative leadership and school improvement: Understanding the impact on school capacity and student learning. *School Leadership & Management, 30*, 95–110.
Hallinger, P., & Heck, R. H. (2011). Exploring the journey of school improvement: Classifying and analyzing patterns of change in school improvement processes and learning outcomes. *School Effectiveness and School Improvement, 22*, 1–27.
Hargreaves, A., & Fullan, M. (1998). *What's worth fighting for out there*. New York, NY: Teachers College Press.
Hargreaves, A., & Macmillan, R. (1995). The balkanization of secondary school teaching. In L. S. Siskin & J. W. Little (Eds.), *The subjects in question* (pp. 141–171). New York, NY: Teachers College Press.
Hargreaves, D. H. (1995). School culture, school effectiveness and school improvement. *School Effectiveness and School Improvement, 6*, 23–46.
Hargreaves, D. H., & Hopkins, D. (1994). *Development planning for school improvement*. New York, NY: Cassell.
Harris, A. (1998). Improving ineffective departments in secondary schools: Strategies for change and development. *Educational Management Administration & Leadership, 26*, 269–278.
Harris, A., Jamieson, I., & Russ, J. (1995). A study of "effective" departments in secondary schools. *School Organisation, 15*, 283–299.
Hattie, J. A. C. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.
Hawley W., & Rosenholtz, S. (Eds.). (1984). Good schools: What research says about improving school achievement [Special issue]. *Peabody Journal of Education, 61*(4).
Hills, J. R. (1983). Interpreting stanine scores. *Educational Measurement: Issues and Practice, 2*(3), 18–27.
Huberman, M., & Miles, M. (1984). *Innovation up close: How school improvement works*. New York, NY: Plenum.
Jackson, D. S. (2000). The school improvement journey: Perspectives on leadership. *School Leadership & Management, 20*, 61–78.
Ko, J. Y. C., Hallinger, P., & Walker, A. D. (2012). Exploring school Improvement in Hong Kong secondary schools. *Peabody Journal of Education, 87*, 216–234.
Lee, J. C.-K., Levin, H., & Soler, P. (2005). Accelerated schools for quality education: A Hong Kong perspective. *The Urban Review, 37*, 63–81.
Levin, H. M. (1987). Accelerated schools for disadvantaged students. *Educational Leadership, 44*(6), 19–21.
Louis, K. S., & Miles, M. B. (1990). *Improving the urban high school: What works and why*. New York, NY: Teachers College Press.
McLaughlin, M. W., & Talbert, J. E. (2001). *Professional communities and the work of high school teaching*. Chicago, IL: University of Chicago Press.
McLaughlin, M. W., & Talbert, J. E. (2006). *Building school-based teacher learning communities: Professional strategies*. New York, NY: Teachers College Press.

Mulford, B., & Silins, H. (2003). Leadership for organisational learning and improved student outcomes – What do we know? *Cambridge Journal of Education, 33*, 175–195.
Mulford, B., & Silins, H. (2009). Revised models and conceptualization of successful school principalship in Tasmania. In B. Mulford & B. Edmunds (Eds.), *Successful school principalship in Tasmania* (pp. 157–183). Launceston: Faculty of Education, University of Tasmania.
Murphy, J. (2013). The architecture of school improvement. *Journal of Educational Administration, 51*, 252–263.
Potter, D., Reynolds, D., & Chapman, C. (2002). School improvement for schools facing challenging circumstances: A review of research and practice. *School Leadership & Management, 22*, 243–256.
Purkey, S. C., & Smith, M. S. (1983). Effective schools: A review. *The Elementary School Journal, 83*, 426–452.
Reynolds, D. (2010). *Failure-free education? The past, present and future of school effectiveness and school improvement*. London: Routledge.
Reynolds, D., Hopkins, D., Potter, D., & Chapman, C. (n.d.). *School improvement for schools facing challenging circumstances: A review of research and practice*. Unpublished paper, University of Exeter, Exeter, UK.
Reynolds, D., Stringfield, S., & Schaffer, E. C. (2006). The high reliability schools project: Some preliminary results and analyses. In A. Harris & J. Chrispeels (Eds.), *Improving schools and educational systems* (pp. 56–76). New York, NY: Routledge.
Reynolds, D., Teddlie, C., Hopkins, D., & Stringfield, S. (2000). Linking school effectiveness and school improvement. In C. Teddlie & D. Reynolds (Eds.), *The international handbook of school effectiveness research* (pp. 206–231). London: Falmer Press.
Rosenholtz, S. J. (1985). Effective schools: Interpreting the evidence. *American Journal of Education, 93*, 352–388.
Rosenholtz, S. J. (1989). *Teachers' workplace: The social organization of schools*. New York, NY: Longman.
Rutter, M., Maughan, B., Mortimore, P., & Ouston, J. (1979). *Fifteen thousand hours: Secondary schools and their effects on children*. Cambridge, MA: Harvard University Press.
Sammons, P., Thomas, S., & Mortimore, P. (1997). *Forging links: Effective schools and effective departments*. London: Paul Chapman.
Saphier, J., & King, M. (1985). Good seeds grow in strong cultures. *Educational Leadership, 42*(6), 67–74.
Sarason, S. (1982). *The culture of the school and the problem of change* (2nd ed.). Boston, MA: Allyn & Bacon.
Schaffer, E., Reynolds, D., & Stringfield, S. (2012). Sustaining turnaround at the school and district Levels: The high reliability schools project at Sandfields Secondary School. *Journal of Education for Students Placed at Risk, 17*, 108–127.
Scheerens, J. (1990). School effectiveness research and the development of process indicators of school functioning. *School effectiveness and school improvement, 1*, 61–80.
Scheerens, J., & Creemers, B. P. M. (1989). Conceptualizing school effectiveness. *International Journal of Educational Research, 13*, 691–706.
Siskin, L. (1997). The challenge of leadership in comprehensive high schools: School vision and departmental divisions. *Educational Administration Quarterly, 33*, 604–623.
Siskin, L. S. (1991). Departments as different worlds: Subject subcultures in secondary schools. *Educational Administration Quarterly, 27*, 134–160.
Slavin, R. E., & Madden, N. A. (2006). *Success for All/Roots & Wings: 2006 summary of research on achievement outcomes*. Baltimore, MD: Johns Hopkins University, Center for Research and Reform in Education.
Sleegers, P., Geijsel, F., & Van den Berg, R. (2002). Conditions fostering educational change. In K. Leithwood & P. Hallinger (Eds.), *Second international handbook of educational leadership and administration* (pp. 75–102). Dordrecht: Kluwer Academic.
Stoll, L., & Fink, D. (1996). *Changing our schools: Linking school effectiveness and school improvement*. Buckingham: Open University Press.
Stringfield, S., Reynolds, D., & Schaffer, E. (2008a). Improving secondary students' academic achievement through a focus on reform reliability: 4- and 9-year findings from the High Reliability Schools project. *School Effectiveness and School Improvement, 19*, 409–428.

Stringfield, S., Reynolds, D., & Schaffer, E. (2008b). Towards highly reliable, high quality public schooling. In MCREL (Ed.), *High reliability organizations in education* (pp. 6–23). Denver, CO: MCREL.

Stringfield, S., Reynolds, D., & Schaffer, E. (2012). Making best practice standard – and lasting. *Phi Delta Kappan, 94*(1), 45–50.

Thoonen, E. E. J., Sleegers, P. J. C., Oort, F. J., & Peetsma, T. T. D. (2012). Building school-wide capacity for improvement: The role of leadership, school organizational conditions, and teacher factors. *School Effectiveness and School Improvement, 23*, 441–460.

Walker, A., & Ko, J. (2011). Principal leadership in an era of accountability: A perspective from the Hong Kong context. *School Leadership & Management, 31*, 369–392.

Willms, J. (1992), *Monitoring school performance: A guide for educators*. London: Falmer Press.