PITFALLS AND SUCCESSES OF PILOT AND LARGE-SCALE LITERACY INTERVENTIONS AT EARLY GRADE,

A LITERATURE REVIEW

Author: Dunston L. Kwayumba, Catholic University of Eastern Africa, Nairobi Kenya.
P.O Box, 100185, 00101 Nairobi; kwayumbad@yahoo.com; +254721304502

Co-author Pascal Wambiya, Mwenge Catholic University, Moshi Tanzania
P.O Box 1226, Moshi, Tanzania; omollo7@yahoo.com ; +255763545935

Co-author Clement Majawa, Catholic University of Eastern Africa, Nairobi Kenya. P.O Box, 62157, 00200 Nairobi; clemajawa@yahoo.com ; +254722659020

Corresponding author: Dunston L. Kwayumba, Catholic University of Eastern Africa, P.O Box, 100185, 00101 Nairobi; kwayumbad@yahoo.com; +254721304502

Abstract

After improved school enrollment especially at primary schools, government across the world and particularly in Africa are now focused on the quality of education. At early grade level, these efforts have been channeled towards improving literacy whose poor mastery has led to poor performance in learning outcomes. This paper reviews some of literacy interventions undertaken at pilot level and at large-scale levels. The review shows that most interventions are successful at pilot but not at large scale. This is majorly contributed by lack of proper planning and inadequate resource mobilization. The paper recommends for the need to take into consideration the context within programs are implemented in during planning in order to realize the success of interventions at large scale.

Key words – Literacy, interventions, pilot, reading, large-scale

1.0 Introduction

Global efforts to have all children access education through the Education for All (EFA) resulted in increased enrollment into primary schools in most parts of the world, in Africa and in Kenya. This enrollment has unfortunately not always followed by the appropriate measures to ensure quality education such as availability of teaching and learning materials, number of teachers available per pupil (Gove et al., 2017) and in many cases classrooms and other physical resources. As a result, questions arise as to whether children are really learning while in school. For example a study undertaken in Kenya showed that children in grade three were not able to read as expected (Uwezo Kenya, 2011). According to Hurry, 2012, schooling should at least produce confident readers and writers who will be able to learn from the school curriculum and manage other literacy expectations in their adult life. Why are children not able to read? In Africa, deficits have been found in the quality and availablity of teaching and learning materials, professional development; assessments required to monitor student performance; and reading instruction (Gove et al., 2017). Thus African governments have partnered with different stakeholders and donors to address these specific issues, to improve the quality of learning, more specifically, reading and numeracy.

The last few decades have seen increased efforts by governments, educators and the international development community to move interventions from small scale and one- off projects. In this regard, focus has shifted to transitioning successful pilot projects to large scale implementation of educational programs with the chief goal of achieving sustainable impact at scale. However questions still remain about the effectiveness of programs when implemented at scale (Gove, Korda & Piper, 2017). It seems that the urgent need for resolution and evidence of action, has resulted in the designing of educational innovations, testing their efficacy, assessing

their effectiveness usually at small scale and sometimes quickly adopting them throughout the system without adequate planning for contextual variations (Serdyukov, 2017).

This shift revives the longstanding debate on whether interventions can be as effective at scale as they are at pilot, globally and in developing countries. There is limited recent literature comparing the effectiveness of interventions when scaled up, to their effectiveness during pilot. Considering the amount of resources that go into scaling up an educational intervention, it is no wonder that researchers, educators and donors alike, are concerned about the impact of interventions programs brought to scale. This is especially so in developing countries where such programs seem to be successful at small scale and when they are managed privately or by donors , yet not so successful when expanded sytemwide to whole countries.There is need for more conversations on the conditions under which research based interventions have been able to succeed or fail in order to better inform practice and policy where there are challenges. The next section presents a review of impact studies of literacy interventions at pilot and large scale with a view to add to the present conversations scaling up of educational interventions in developing countries. The following questions will be explored: First, whether early grade literacy interventions are as successful when scaled up as they were during pilot? Secondly, what factors influence the potential for effectiveness of a scaled up literacy intervention.

In this paper scaling up refers to the replicating, expanding and sustaining of successful literacy interventions to cover a larger geographical area with a view to reach a greater number of students (Schneider & McDonald, 2006). Scaled up interventions can be national or near national in geographical scope. According to Cooley and Linn (2014), a pilot program or project entails a trial of a new idea, model or approach with limited impact, monitoring and evaluation of the project, where new knowledge is realised and systematically disseminated. This paper presents the patterns of gains from literacy interventions that have evolved from small scale to implementation in multiple sites globally and in Kenya, it then discusses some challenges and factors that impact on the scaling up of literacy interventions particulalrly in Africa, and concludes with recommendation of what needs to be done to grow the gains in scaled up literacy interventions.

2.0 Literature review

Reports of widespread inability to read among students who have gone through years of years of schooling, and the increased demand for accountability from stakeholders has compelled governments and educators to look for urgent solutions to the reading problem. As a result,

education policy has supported a number of activities designed to improve these reading scores, among them, small scale literacy interventions that have been found to be successful have been adopted, funded and simultaneously rolled-out for implementation to wider contexts either nationally, or near nationally and some across countries (Cove et al., 2017). Adopting such systemwide early grade interventions is believed to reliably replicate the effects seen in small scale studies.

The Reading Recovery is an intensive reading program adopted widely by many schools in the United States to address the reading problem. The RR program, originally developed by Marie Clay in New Zealand is widely used in New Zealand and has received wide attention in the United States. Like many other widely used interventions, this program was first piloted with a small group before it expanded to other states. Two colleagues, Marie Clay and Barbara Watson, trained a number of teachers to use the program and also trained some faculty members from Ohio State University to serve as key trainers. Since then the program expanded to other countries and throughout New Zealand and the United States.

The RR program has been widely evaluated for effectiveness in the various countries where it is used in large scale (Chamberlin, 2015; Reynolds & Wheldall, 2007; Sirinides, Gray & May, 2018). For example, a randomised field experiment by May et al., (2013) found effect sizes larger than 0.40 SD for the RR program. Effect sizes considered moderate by Cohen (1992) but considered adequate for determining an effective educational intervention by Hattie (2012). The study by May and colleagues used rigorous experimental and quasi experimental designs to estimate impact by the program. Another rigorous impact study of the Reading Recovery program was carried out in the United Kingdom by Hurry (2012) and found effect sizes of up to 0.39 SD. The study by Hurry (2012) found these effects of the RR long after the learners had completed the intervention.

Among the few studies to report negative results is the one by Chapman and Tunmer (2015) which argued that the program was not as effective as widely believed. Their design however, has been criticised as not being rigourous enough to merit their results and as such the findings should be interpreted with caution (Schwartz, Hobsbaum, Briggs & Scull, 2009). Schwartz and colleagues opine that there are enough high quality studies that demonstrate causal validity of the RR program. Additionally, the studies reviewed by Schwartz and colleagues demonstrate the rigour during implementation where student data is collected at entry and at exit; there is continous monitoring and evaluation and problems are identified early and addressed. While Schmidt and Gregory (2005)

believe that large scale literacy interventions can lead to positive results, the system wide interventions have not always led to the expected magnitude.

The Success for All (SFA) was another widely implemented program in the United States. The three year trial of the intervention by the non - governmental organisation supporting the program realised significant impact, enough to influence a scale up of the program. The program began in Baltimore where it was piloted in one public school. The emphasis of SFA right from the outset was to have rigorous evaluation of its outcomes by using individualised reading assessments (Slavin et al., 1994). Slavin and colleagues acknowledged that limitations in funding impacted in schools ability to assess their students in the recommended manner. Important is the built in design to cater for local circumstances and variations across contexts. The early evaluations of the SFA before wide expansions yielded effect sizes of 0.31 SD to 1.77 in Baltimore, up to1.67 SD in philadelphia. other districts are recorded (*see slavin et al., 1994*).

After it scale up, the program did not record as huge effects as it had done in its small scale implementation evaluations (Borman,et al., 2007). Quint et al.,(2015) cited the lack of sufficient teaching resources and big variations in grouping students in the various schools. The prevailing economic recession at the time of its scale up led to large cuts in public spending. Additionally, teachers, who are key stakeholders in educational intervention, did not give their full support to the program because they felt that the program undermined their professional knowledge and skills gained from years of experience in classroom practice.

A three year whole school early grade literacy pilot program Pact of Literacy at the Right Age(*Pacto pela Alfabetização na Idade Certa*[PAIC]), was implemented in Brazil to test its efficacy in dealing with the literacy problem in the country. An evaluation of the program after the three years showed a significant decrease of non readers from 14% to 6%. Costa and Carnoy (2015) carried out an impact evaluation of the scaled up version, the National Pact of Literacy at the Right Age (*Pacto Nacional pela Alfabetizacao na Idade Certa, [PNAIC]).* They found a positive effect on student achievement in Mathematics and portuguese but did not find significant differences in the overall learning achievement of 6-8 year old children from public schools in Brazil.

Jordan and Egypt rolled out a nation wide early grade literacy program that came after the successful small scale trial in 41 schools and 166 primary schools respectively (Gove, Brombacher & Ward-Brent, 2017). Another early grade literacy program was piloted in several African and Asian countries including Malawi, Rwanda, Burundi, Ethopia, Bangladesh, Indonesia and the Philippines. The program was later expanded to wider implementation across 10 countries. An early

evaluation of the program's impact by Guajardo and Ochoa (2012) revealed effect sizes of 0.27 to 1.21, which could be considered moderate to large (Hattie, 2012). A more specific analysis of the impact of the literacy boost program in Ethiopia by Guajardo, Wubeshet, McVitalis, Ochoa and Dowd (2012), which is one of the countries implimenting the program managed by Save the Children, shows a variation in contextual factors. Factors such as availability of physical resources within the school including presence or absence of electricity, natural classroom light, average tenure of teachers in schools, proximity of schools to children's homes accounted for higher or lower test scores. A cross country analysis of the literacy boost program by Dowd, Friedlander, Guajardo, Mann and Pisani (2013) showed just how these contextual factors can impact the outcome of literacy interventions. While there were significant gains in learning outcomes of effect sizes 0.25 to 0.75 in Ethiopia, Malawi, Nepal, Pakistan and Zimbabwe, the impacts of the literacy program were not as easy to compute in Bangladesh and Mozambique. This was due to high attrition in Mozambique and difficulties with community implimentation in Bangladesh.

Similar challenges are observed in the EGRA Plus in Liberia and another literacy intervention in Rwanda as recorded by DeStefano and Healey (2016). The literacy program in Rwanda, produced an effect size of *0.55 SD* when the program was implemented in small scale covering 90 schools but reduced to *0.19 SD* after scale up. In this particular program, there were failures in delivering learning and teaching material to schools in addition to logistical challenges in the teacher support.These failures were mainly due to inability by the government wing in planning and human resourcing. The Liberia Teacher Training Program 2 (LTTP2), had difficulties with implementation in 1000 schools after a successful pilot in 120 schools. The reduced impact was linked to limited human resource and challenges with physical infrastructure which caused delays in the delivery of learning materials to schools.

Kenya has implemented a number of educational interventions in the recent past. These interventions were based on evidence of success in small scale trials only to produce smaller impacts or fail altogether when rolled out across the nation. Early on, Kenya found short comings with the existing pedagogical practices and it was hypothesised that in- service teacher training and provision of teaching and learning materials would change the way students and teachers interacted in the classroom, to improve quality of primary school learning. Towards this end, the Governement of Kenya through the Ministry of Education launched the Primary school Management Project (PRISM) to train head teachers with the aim of curriculum and administrative reforms. The apparent success of this program, carried out in public primary schools in Kenya led to the launch of the

School based Teacher Development (SbTD) program under the SPRED III which ran from 2001-2005. SPRED I and II had focused on provision of teaching and learning materials and head teachers training respectively.

The SbTD program was directed towards classroom teachers and had the goal to provide in-service teacher professional development in reading and mathematics. This was to be cascaded to other teachers by utilising Key Resource Teachers. Within this same time, the Ministry of Education (MoE) introduced another program Instructional Material Programme (IMP) aimed at improving classroom teaching and learning by providing instructional materials to promote reading skills in primary school children. An evaluation of the program by Hardman et al., (2009) shows that the SbTD program successfull met its training target, which resulted in the selection of Key Resource Teachers (KRTs) to spearhead the school based professional development in their schools. Their study used a mixed method approach sampled 12 districts and interviewed teachers, pupils, headteachers and school management committees. The study found that the program had improved classroom practices and student-teacher interraction. The cascade model of teacher training was found to have less impact than anticipated and the poor performance was due to a heavy teacher workload. A study using a descriptive design revealed that the program faced challenges that ranged from inadequacy of the cascade model of training; teachers' unmet expectations and inadequate numbers of educators (Gathumbi, Mungai & Hintze, 2013).

The Strengthening of Mathematics and Science in Secondary Education (SMASSE) intervention is another such program that did not adequately achieve its goals. After a successful pilot in 15 schools in 9 out of 72 districts in Kenya, the program was quickly expanded to other public schools nationally (Mwangi & Mugambi, 2013; Waititu & Orado, 2009). According to Waititu and Orado (2009), the program successfully trained up to 4000 teachers and this was attributed to the favourable political support at the time and ownership by the MOE and the school teachers. There was also effective management of the program which led to the expansion of the program to primary schools from 2009. However, this program also used the cascade method for teacher training which faced challenges of distortion and unnecessary long period before knowledge could trickle down to learners. The persistent problem of unmet teacher expectations and teacher perceptions that the program assumed or overlooked their professional knowledge, led to compromise in that teachers were not willing to change their classroom practices in favor of the methods taught by SMASSE. Additionally, in both of these programs, learning materials were

either not available, were limited and were often not delivered to the schools on time and the cascade method of training failed in this case (Gathumbi, Mungai &Hintze, 2013).

Currently, Kenya is implimenting the Tusome Early Grade literacy intervention, a program that came after a successful trial of the Primary Maths and Reading (PRIMR) initative PRIMR was designed and implemented by the MoE with technical support from RTI international and funding from the USAID. The pilot of the program was carried out in 547 urban and peri-urban schools in five counties in Kenya. In due consideration of the variation in contextual factors and potential challenges at scale up, the Tusome Early Grade literacy intervention made room for possible decrease in effect size (Piper, Zuilkowski and Mugenda, 2014)

Discussion

The review showed that literacy interventions did not always return positive or large impacts at scale-up, even in places where several trials of the interventions showed large impacts. The key determining factor was the differentiated environments and contexts within which the programs were implemented. Additionally, from the review, it seems that nations and systems that are well resourced were more likely to realise positive impacts on scaled up literacy interventions when compared to the less resourced systems which struggle with various challenges and political and leadership support issues.

Schwartz et al., (2009) argued that well resourced sytems and those with adequate national support are more likely to register gains from educational interventions. Major factors emerging as impediments to the effective implementation include lack of teaching and learning materials, physical classrooms and other school resources, teacher attitudes towards the literacy interventions, political factors within the school or country. Stein et al., (2008), found that technical support at the intervention site had major impacts on the effects of an intervention on learners reading achievement. The willingness and ability of teachers to adopt instructional practices recommended in an intervention was a mediating factor to the effectiveness of an intervention.

The challenges of scaling up educational interventions have been considered historically. It is encouraging to observe that even with a strong knowledge base of scientific research on scaling up, disciplines that began the scale up conversations long before education are still learning on how to effectively scale up in a variety of contexts and with people other than the initial developers of an innovation (Schneider & McDonald, 2006). When thinking about scaling up and multi-site implimentation, attention is drawn to fidelity of implimentation and as argued by Sloane (2005) it

seems unlikely that the implimentation will be carried out with similarity across the various sites when compared to implementation in smaller units. This is mainly because, rolling out systemwide programs happens in the context of several factors including educational policies, varied expectations of the different stakeholders, varied institutional conditions, sometimes limited resources. Cooley et al., (2014) termed these factors as 'scaling up spaces' as essential factors that may influence the success of a scaled up intervention. Cooley and colleagues identified institutional, policy, financial, political, partnership and learning spaces as critical factors which if well handled would lead to the success of a scaled up innovation.

More specifically, as argued by Schneider and McDonald (2006), 'prototypes' that work for one population may not always work for another and their is need for close managment, monitoring and evaluation and well defined measurement tools, if small scale studies are to succeed as large scale interventions. One matter they raise is the communication and dissemination of results of efficacy studies with practitioners and not just national decision makers. As noted during this review, some teachers felt that their years of expereience and their knowldge as practitioners was ignored. Perhaps if empirical findings are shared with such practitioners during workshops, conferences, teacher education and other in-service training opportunities, then they may have opportunity to participate in the drafting of the final program that goes to scale. Additionally, with credible findings, arising from rigorous objective standards of research, the implementing stakeholders might be more reassured.

Furthermore, the pilot programs particularly in developing countries are often managed by well resourced and well organised non- governmental organisations. Since in these organisations, projects are sytematically monitored and problems, mistakes are lessons learnt and are used to improve the programs (Bold et al., 2012). Scale up requires partnerships with the national government through ministries of education, with community members, teachers and school leadership, all of which sometimes miss the characteristics of careful, systematic organisation and committment of NGO's and donors. Slavin (2008) however warns that sometimes the results of pilot programs are overstated or biased especially when reporting organisations are commercial and need positive publications of the interventions.

Finally, it was noted that there were limited evaluation studies of interventions in Africa. In evaluation of both pilot and national interventions, not all studies reported an assessment of implementation fidelity, methods of assessing reading were not standard thus introducing a limitation in the comparison of pilot and national programs. This was a problem observed also by

Brenner and Hiebert (2010). They cautioned that the lack of consistency in reading assessment tools for instance in some of the studies would need to be addressed in order to assure validity of the evaluations.

3.0 Conclusion and recommendations

It is evident from the various evaluation studies that large-scale implementation of research-based interventions is bound to yield differentiated results stemming from the varied settings in which the interventions are implemented. Emerging also is the need for definition of a local benchmark for the research designs that can be used to assess literacy interventions. There is also need for a general guideline on what kind of approaches or models of educational interventions can be effective in the African and Kenya context and conditions.

There is need to undertake sufficient planning before scaling up projects. As mentioned, pilot projects are often undertaken on a small scale and the planning and preparations activities required before their commencement might not be as detailed as those required at for large-scale. The proportion of project players increase in a direct proportion with the increment in the project size. This calls for adequate planning in order to meet the needs of all the players. Other aspects that need to be planned in adequate include resources for the project. For example, a project that provides a literacy interventions with various components including teaching and learning materials, teacher training and instructional support should plan adequately for the astronomical increment in the pupils that will receive instructional materials, teachers that will receive training and the instructional support team. The coverage area should also be taken into consideration in the plans.

There is need to acquire adequate resources as projects scale-up. As mentioned, the amount of resources required proportionately increase with the scope of studies. Some of the projects reviewed were not successful because it was assummed that there would be cost-sharing. Parties involved might not have resources for cost-sharing or the resources might not be sufficient. Therefore in order for projects to be suistanable at large scale, there is need for implementors to mobilize enough resources to enable the projects to run as they did at pilot.

Some of the projects succeed at pilot because of the scientific explanation such as the *Harthorne effect* (Best & Kahn, 2011). Therefore during pilot, will most likely be determined to show that the project works. The levels of motivation will be high but unsuistanable. Therefore in order to harness the effect of project at scale-up, there is need to put in place aspects of motivation.

Motivation can be monetory (in terms of realistic allowances) or non-monetory (in terms of recognition through certification).

References

Best, W. B., & Kahn, V. J. (2011). *Research in Education* (10th ed.). New Delhi: PHI Learning Private Limited.

Bold, T., Kimenyi, M., Mwabu, G., Ng'ang'a, A., & Sandefur, J. (2012). Interventions & institutions experimental evidence on scaling up education reforms in Kenya. *Preliminary draft. Available at http://www. iies. su. se/polopoly_fs/1.101632, 13481*, 37980.

Borman, G. D., Slavin, R. E., Cheung, A. C., Chamberlain, A. M., Madden, N. A., & Chambers, B. (2007). Final reading outcomes of the national randomized field trial of Success for All. *American Educational Research Journal*, *44*(3), 701-731.

Brenner, D., & Hiebert, E. H. (2010). If I follow the teachers' editions, isn't that enough? Analyzing reading volume in six core reading programs. *The Elementary School Journal*, *110*(3), 347-363.

Chamberlin, R. M. (2015). *Reading Recovery Methodology as Special Education Literacy Intervention* (Doctoral dissertation).

Cohen, J. (1992). A power primer. *Psychological Bulletin*, pp. 155-159.

Cooley, L., & Linn, J. F. (2014). Taking innovations to scale: methods, applications and lessons. *Washington DC: Results for Development Institute*.

Costa, Leandro Oliveira, and Martin Carnoy. 2015. "The Effectiveness of an early grade literacy intervention on the cognitive achievement of Brazilian students." *Educational Evaluation and Policy Analysis 37 (4)*, 567–90.

DeStefano, J., & Healey, F. (2016). *Scale-up of early grade reading programs*. USAID/Asia Bureau

Dowd, A. J., Friedlander, E., Guajardo, J., Mann, N., & Pisani, L. (2013). Literacy Boost cross country analysis results. *Save the Children, Washington, DC*.

Gathumbi, A. W., Mungai, N. J., & Hintze, D. L. (2013). Towards comprehensive professional development of teachers: The case of Kenya. International Journal of Process Education, 5, 3–14.

Gove, A., Korda Poole, M., & Piper, B. (2017). Designing for scale: Reflections on rolling out reading improvement in Kenya and Liberia. *New directions for child and adolescent development, 2017*(155), 77-95.

Gove, A., Brunette, T., Bulat, J., Carrol, B., Henny, C., Macon, W., Nderu, E., & Sitabkhan, Y. (2017). Assessing the impact of early learning programs in Africa. In Kenneth R. Pugh, Peggy McCardle, & Annie Stutzman (Eds.), Global Approaches to Early Learning Research and Practice. New Directions for Child and Adolescent Development. 158, 25–41.

Guajardo, J., & Ochoa, C. (2012). *Literacy Boost: Lessons from assessment and community*. Presentation at the All Children Reading Workshop, July.

Guajardo, J., Wubeshet, F., McVitalis, D. O., Ochoa, C., & Dowd, A. J. (2012). Literacy Boost Ethiopia: Baseline report. *Save the Children, Washington, DC*.

Hattie, J. (2012) Visible Learning for Teachers: Maximizing Impact on Learning. Londno: Routledge.

Hardman, F., Abd-Kadir, J., Agg, C., Migwi, J., Ndambuku, J., & Smith, F. (2009). Changing pedagogical practice in Kenyan primary schools: the impact of school-based training. *Comparative Education, 45*(1), 65-86.

Hurry, J. (2012). The impact of Reading Recovery five years after intervention. *report for the Every Child a Reader Trust, London: Institute of Education, University of London. http://www. ioe. ac. Uk/about/documents/Hurry_London_follow_up_2012_Report_december_12. pdf*.

Mwangi, N. I., & Mugambi, M. (2013). Evalaution of strenghtening of mathematics and science in secondary education (SMASSE) Program. A case study of Murang'a South District, Kenya. *International Journal of Education Learning and Development, 1,* 46–60.

Piper, B., Zuilkowski, S. S., & Mugenda, A. (2014). Improving reading outcomes in Kenya: First-year effects of the PRIMR Initiative. *International Journal of Educational Development, 37,* 11-21.

Quint, J, Balu, R., DeLaurentis, M., Rappaport, S., Smith, T., & Zhu, P. (2015). *The Success for All model of school reform: Findings from the Investing in Innovation (i3) scale-up*. New York: MDRC.

Reynolds, M., & Wheldall, K. (2007). Reading Recovery 20 years down the track: Looking forward, looking back. *International Journal of Disability, Development and Education*, *54*(2), 199-223.

Schmitt, M. C., & Gregory, A. E. (2005). The impact of an early literacy intervention: Where are the children now? Literacy Teaching and learning: An International Journal of Early Reading and Writing, 10(1), 1-20.

Schneider, B., & McDonald, S. K. (Eds.). (2006). *Scale-up in education: Ideas in principle* (Vol. 1). Rowman & Littlefield Publishers.

Schwartz, R. M., Hobsbaum, A., Briggs, C., & Scull, J. (2009). Reading Recovery and Evidence-based Practice: A response to Reynolds and Wheldall. *International Journal of Disability, Development and Education*, *56*(1), 5-15.

Slavin, R. E., Madden, N. A., Dolan, L. J., Wasik, B. A., Ross, S. M., & Smith, L. J. (1994). 'Whenever and Wherever We Choose': The Replication of'Success for All'. *The Phi Delta Kappan*, *75*(8), 639-647.

Sirinides, P., Gray, A., & May, H. (2018). The Impacts of Reading Recovery at scale: Results from the 4-Year i3 external evaluation. *Educational Evaluation and Policy Analysis*, 0162373718764828.

Slavin, R. E. (2008). Perspectives on evidence-based research in education—What works? Issues in synthesizing educational program evaluations. *Educational researcher*, *37*(1), 5-14.

Sloane, F. C. (2005). The scaling of reading interventions: Building multilevel insight. *Reading Research Quarterly*, *40*(3), 361-366.

Stein, M. L., Berends, M., Fuchs, D., McMaster, K., Sáenz, L., Yen, L., ... & Compton, D. L. (2008). Scaling up an early reading program: Relationships among teacher support, fidelity of implementation, and student performance across different sites and years. *Educational Evaluation and Policy Analysis*, *30*(4), 368-388.

Uwezo Kenya (2011). Are our children learning. *Annual Learning Assessment Report: Uwezo Kenya, Nairobi*.