KU LEUVEN H:VA
RESEARCH INSTITUTE FOR WORK AND SOCIETY

ON THE CONTRIBUTIONS OF REMEDIAL TEACHING TO SOCIO-EMOTIONAL LEARNING OF CHILDREN IN THE PRIMARY EDUCATION SYSTEM OF ZAMBIA

Sofie Cabus, Sharon Schroen, Michael Phiri, Mooya Haatembo, Stefaan Vandewalle & Nico Vromant

Published by
KU Leuven
HIVA - RESEARCH INSTITUTE FOR WORK AND SOCIETY
Parkstraat 47 box 5300, 3000 LEUVEN, Belgium
hiva@kuleuven.be
http://hiva.kuleuven.be

Declarations

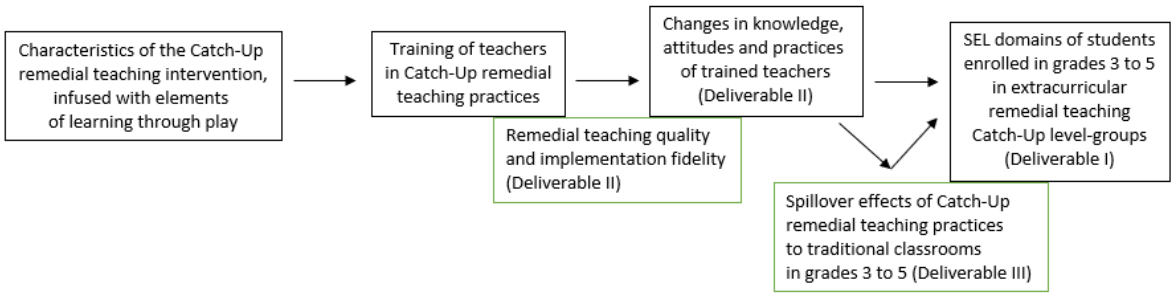
1) **Funding**
   The LEGO Foundation.
2) **Conflicts of interest**
   The authors declare that they have no relevant material or financial interests that relate to the research described in this paper.
3) **Availability of data and material**
   The Centre for the Promotion of Literacy in Sub-Saharan Africa developed a survey and collected data using this survey. Data to replicate results are available.
4) **Ethics approval**
   An application for research ethical clearance was submitted to the ethical review board of the University of Zambia and was approved in October 2021.
5) **Consent to participate**
   The research posed minimal risk to the participants. Informed consent was obtained for all survey respondents prior to the commencement of the data collection.
6) **Consent for publication**
   All stakeholders involved in this study, including the authors, gave informed consent to publish this manuscript. This article only presents group-level information and refrains from including personal identifiable information.

Contents

Executive summary

Besides demonstrated impacts on students' foundational learning outcomes (Asio & Jimenez, 2020; Panayiotou et al, 2019; Selvarajan, 2022; Walker eta al, 2021), remedial teaching interventions are believed to positively influence students' social and emotional learning (SEL). SEL has been shown to be associated with increases in students' learning outcomes (Puerta et al., 2016; Panayiotou et al, 2019; Taylor et al., 2017; Weissberg et al., 2007). However, limited information exists on the relationship between remedial interventions and SEL, specifically for low-and-middle income countries. The current study aims to contribute to this existing gap by investigating the impact of the Catch-Up remedial teaching program in Zambia on teachers' knowledge, attitudes, and use of Catch-Up promoted classroom practices and, eventually, students' SEL outcomes. The research rationale is captured in Figure 1. The figure was inspired by the framework of Desimone (2009), that visualizes the 'trickling down effects' of teacher professional development to students.

Figure 1: Research rationale with reference to the deliverables



Source: This figure is partially based on the theoretical framework of Desimone (2009).

The first box to the left in Figure 1 deals with the characteristics of the Catch-Up remedial teaching intervention, followed by the effective training of teachers. Catch-Up is a remedial teaching program implemented in Zambia, which aims to promote foundational numeracy and literacy skills among children in grades 3-5 of primary education. The Catch-Up program is an adapted version of the Teaching at the Right Level (TaRL) approach, which has been adjusted to fit the Zambian education context. It integrates several elements of Learning through Play (LtP). These LtP elements are based on the LEGO Foundation's characteristics of LtP, contextualized in the Zambian education setting by the Ministry of General Education (MoGE) in the 7Cs framework. This framework recognizes 7 components of play in the Zambian setting (concrete, cheerful, captivating, collaborative, creative, challenging, and connected).

We explore the impacts of the Catch-Up remedial teaching intervention on student outcomes regarding socioemotional learing (Deliverable I); and (Deliverable II) remedial teaching quality and implementation fidelity and the knowledge, attitudes and practices of trained teachers. Furthermore, we also explored the potential impacts of training teachers in Catch-Up on regular teaching practices (Deliverable III). Mapping the impacts of Catch-Up, we employed a mixed-method approach to investigate the impact of the Catch-Up program on student and teacher outcomes. Data was collected

at both baseline (February-March 2022, 3 weeks after the start of Catch-Up implementation) and endline (September-November 2022, 7-9 months after the start of Catch-Up implementation). To determine the impact of the program on learners' SEL outcomes, 2,606 learners from 95 schools in Lusaka province (treatment sites) and Central province (control sites) were administered an adapted version of the International Social Emotional Learning Assessment tool (ISELA).

Besides measuring outcomes at the student level, the current study also deemed it necessary to research what happens at the teacher level. Without a proper understanding of how teachers are understanding and implementing the Catch-Up program, it becomes difficult to appraise student outcomes. Since the Catch-Up program is infused with LtP elements, teacher tools focus on measuring their opinions, understanding, and use of LtP, besides other Catch-Up promoted practices. One teacher per school was selected to participate. For treatment schools, the participating teacher was always a Catch-Up teacher. There were 4 different tools measuring teacher outcomes:

- *Implementation fidelity observations*: these observations were conducted at treatment schools only. Trained enumerators observed teachers conducting Catch-Up classes and filled out a checklist measuring adherence to general Catch-Up guidelines. A total of 25 teachers were observed at endline.
- *Teacher surveys*: teachers at both treatment and control schools participated in self-reported surveys, measuring their knowledge, attitudes, and use of Catch-Up promoted classroom practices. A total of 86 teachers filled out this survey at endline.
- *Classroom observations*: regular classes of teachers at both treatment and control schools were observed by trained enumerators. Enumerators filled out a survey measuring teachers' use of Catch-Up promoted classroom practices, with a specific focus on the use of learning through play as discussed in the 7Cs framework. A total of 77 Catch-Up classes were observed at both baseline and endline.
- *Qualitative interviews*: at treatment schools only, 25 teachers and 5 stakeholders were sampled to participate at endline. These interviews sought an in-depth understanding of how the respondents understand the roles of teacher and learner in the classroom, what they consider effective and engaging teaching, their use of LtP elements as discussed in the 7Cs framework in their teaching, and how they view the learning process.

Overall, the Catch-Up program showed impacts on some of the student and teacher outcomes of interests:

- A treatment effects model using a robust-correction approach to missing data in a randomized setting indicated that there were small but significant increases in learners' capacity to understand and empathize with others (Cohen's d = 0.141 *SD*), as well as their ability to navigate and resolve conflicts in social settings (Cohen's d = 0.121 *SD*). There was a significant decrease in learners' scores on the network domain (Cohen's d = -0.083 *SD*), and an absence of significant impacts on the other 3 domains of self-concept, stress management, and perseverance.
- The implementation fidelity tool showed high adherence to Catch-Up promoted guidelines at participating treatment schools. Percentages of adherence ranged from 64% to 88%. For adherence to guidelines measured on a quantitative scale, most of the values were significantly higher than the midpoint of the scale. Adherence to desired class duration and the use of group leaders to lead classroom activities is sufficient but should ideally be improved.
- For teacher surveys, ANOVA analyses investigating the impact of the Catch-Up intervention on teachers' knowledge, opinions, and use of Catch-Up promoted practices showed mostly inconclusive results. However, teachers at treatment schools did score higher than teachers at control schools on items measuring the use of classroom practices that reflected the Cheerful and Collaborative components of the 7 Cs framework. Teachers at treatment

schools also scored higher on items measuring their contribution towards and enjoyment of learners' progress, while teachers at control schools scored higher on items measuring their satisfaction with the compensation and equipment received for their job. Only some of these differences reached statistical significance. Teachers at treatment schools reported positive attitudes towards the Catch-Up program from both them and others in their school environment, with most of the items scoring significantly higher than the midpoint of the scale.

- ANCOVAs were performed to determine the impact of the Catch-Up intervention on the endline scores of teachers on the classroom observation tool, while including their baseline scores as a covariate. Treatment schools scored consistently higher than control schools on items measuring the frequency of use of Catch-Up promoted practices in classrooms, with items measuring the use of playful activities (e.g. songs, dances, games, and stories) gaining statistical significance. For the use of activities reflecting the components of the 7Cs framework, 4 out of 7 components (Concrete, Cheerful, Captivating, and Collaborative) showed higher scores for treatment schools than control schools. For the Concrete component, this difference between treatment and control schools gained statistical significance ($p = .029$), while for the Cheerful component, the difference approached significance ($p = .060$). The Creative component showed that treatment schools scored higher than control schools at endline, however this difference did not reach statistical significance and decreased between baseline and endline. Results for the Challenging and Connected components were inconclusive.
- Qualitative interviews showed that both teachers and stakeholders can describe a wide array of effective teaching methods and engaging classroom activities, that reflect an understanding of how to implement the 7Cs framework and Catch-Up promoted practices in their classrooms. Teachers are specifically focused on teaching methodologies and activities that reflect the Concrete, Cheerful, Collaborative, and Captivating components of the 7Cs framework. The Creative and Connected components are less often quoted in teachers' descriptions of their teaching and practices, and the Challenging component seems absent. Stakeholder answers more often report on these components, which shows they have a thorough understanding of the 7Cs framework and what it entails.

Teacher outcomes showed consistent trends between deliverables but results often lacked statistical significance. While the focus of these deliverables was qualitative, future research should investigate these outcomes with more rigor and larger sample sizes, to see if current trends can be replicated and gain statistical significance. At the same time, it is acknowledged that not all teachers in the schools are engaged in the Catch-Up training which of course impacts the potential sample size. In addition, the teacher measurements largely focused on teachers' use of Catch-Up promoted practices in general classrooms. However, the use of these practices in Catch-Up classes specifically was not extensively measured. This makes it difficult to determine whether the absence of some of the desired teacher outcomes is caused by teachers not implementing these Catch-Up practices at all, or whether this is due to a lack of spillover effects between Catch-Up classes and regular classes. Further research should measure the classroom practices of teachers in Catch-Up classes more extensively, to be able to answer these questions.

While we found small but significant impacts on student and teacher outcomes, there were several limitations to this study that should be considered when interpreting the results. Even though the implementation fidelity observations showed positive results, it is important to note that some of the treatment schools failed to participate in these observations, due to the absence of Catch-Up teachers or Catch-Up classes being organized at the day of data collection. In combination with anecdotal evidence from project staff on difficulties with quality implementation of the program,

implementation fidelity might not be as high at these schools as what was measured in this study. Besides that, there are issues concerning Catch-Up refresher training. Some teachers report not being invited to them, or state they take place later than desired (Busara, 2023). Literature on quality teacher professional development stresses the importance of such follow-ups when development activities take place in a short time frame (which is the case in Catch-Up) to promote intellectual and pedagogical change (Desimone, 2009; Merchie, Tuytens, Devos, & Vanderlinde, 2016). Hence, problems with the adequate implementation of refresher training could negatively impact changes in teachers' knowledge and practices. The timing of the endline should also be taken into consideration when looking at the results of this study. The endline took place only 7-9 months after program implementation, which might have been too short for effects to manifest (Merchie et al., 2018; Popova et al., 2022). In addition, frequent teacher transfers between classes and schools could have also resulted in teachers participating in the endline who had only recently been exposed to Catch-Up training, and thus had had little time to change their teaching practices as a result of the Catch-Up program. Future research should look into the impacts at a later point in time, as well as measure teachers' level of exposure to the Catch-Up program, to be able to statistically control for differences in their experience.

As shown on Figure 1, the current study is divided into 3 different deliverables. These three deliverables are structured in this report as follows. Deliverable I discusses the impact of the Catch-Up program on learners' SEL outcomes. Deliverable II looks at the implementation fidelity in Catch-Up classes and the teachers self-reported surveys on Catch-Up knowledge, opinions, and practices. Deliverable III describes the classroom observations of regular classrooms and qualitative interviews of teachers and stakeholders. Below, a general introduction and a description of the Catch-Up remedial program follows. After that, each deliverable is discussed individually, including the following sections: Methods, Results, Conclusion and discussion.

Deliverable I: Catch-Up and SEL of children

Introduction

Remedial interventions are proven successful in tackling children's backlog in foundational skills (Asio & Jimenez, 2020; Panayiotou et al, 2019; Selvarajan, 2022; Walker eta al, 2021). These interventions have targeted varying foundational skills in primary education including numeracy and literacy. Furthermore, these interventions often include continuous professional development of in-service teachers or trainers to get them familiar with the didactic materials and approach (Merchie et al., 2018; Popova et al., 2022). While previous studies indicated significant effects of remedial teaching on student outcomes (among others, Asio & Jimenez, 2020; Banerjee et al., 2007, 2016; Lakshminarayana et al., 2013; Metsäpelto & Pulkkinen, 2014), there is only limited evidence yet on its contributions to socioemotional learning (SEL). This particularly holds true for low-and-middle-income countries. There are reasons to believe, however, that remedial interventions can contribute to several domains of SEL, whereas increased learning outcomes have been previously associated with SEL (Puerta et al., 2016; Panayiotou et al, 2019; Taylor et al., 2017; Weissberg et al., 2007).

The limited evidence available on low-and-middle-income countries are coming from recently published studies by Aber et al. (2017), Torrente et al. (2015, 2019) and Brown et al. (2021) on classroom climate-targeted SEL programs in the Democratic Republic of Congo and Niger, and by Tubbs Dolan et al. (2022) and Kim et al. (2023) on nonformal remedial support and mindfulness programs in Lebanon. These studies indicate an overall small to moderate impact of remedial programs infused with SEL practices on children's perception of the school environment, such as the supportiveness of teachers and schools. The studies also explored the effectiveness of nonformal after-school remedial teaching versus remedial teaching as part of the formal school curriculum. Aber et al. (2017), Torrente et al. (2015, 2019) find that literacy and numeracy skills improved in the nonformal setting (thus, not the formal school grades), while Brown et al. (2021) find that both children's perceptions on the school environment and literacy and numeracy skills increased in the formal setting. The study of Tubbs Dolan et al. (2022) focused on SEL-infused remedial teaching in a nonformal setting targeting a vulnerable group of Syrian refugee children, and found promising moderate effects of this program on children's perceptions of the school environment, and small effects on foundational numeracy and literacy skills and SEL outcomes.

This paper contributes to the previous literature by focusing on the contributions of a remedial teaching program in a low resource context to increasing SEL. We focus on the effectiveness of a remedial intervention, Catch-Up, that was implemented in the primary education system in Zambia, Lusaka Province, in the years 2021-2023. The Catch-Up (CU) program is based on the remedial Teaching at the Right Level (TaRL) methodology, earlier developed and implemented by Pratham, an Indian NGO, in the Indian primary education system, and tested on its effects on literacy and numeracy (Banerjee et al., 2007, 2016; Vroman et al., 2021). While the program at its core is a remedial teaching program in foundational literacy and numeracy, targeted at children in grades 3 to 5 enrolled in the formal education system, it also uses several elements of learning through play (LtP). LtP is associated with socioemotional development in many studies (Broadhead, 2006; Hamre & Pianta, 2005; Lungu & Matafwali, 2020; Moran, 1971; Parker & Thomsen, 2019; Whitebread et al., 2017; Zosh et al., 2017). These studies give us reasons to believe that a LtP-infused remedial teaching program can boost SEL.

This paper proceeds as follows. In Section 1 | we discuss the Catch-Up remedial teaching program. Section 2 | presents the methods of data collection and the empirical strategy. The results are discussed in Section 3 | . Section 4 | concludes.

1 | Catch-Up remedial teaching program

1.1 From Pilot to Scale

In 2015 Zambia was facing a learning crisis. It had been ranked last in measures of literacy and numeracy by the 2011 Southern and Eastern Africa Consortium for Monitoring Education Quality (SACMEQ), and a 2014 national assessment found that 65% of Zambian grade 2 learners were unable to read a single word in their local language. About 38 percent of grade 5 learners were illiterate and had poor math skills (Zambian Ministry of Education, 2016). In response, the Ministry of General Education (MoGE) looked for ways to address the foundational skills gap. It was particularly keen to adopt and adapt the Teaching at the Right Level (TaRL) approach. J-PAL Africa supported this process by targeting key development actors, sharing evidence on TaRL and discussing how it could be applied in the Zambian context.

MoGE created a working group in 2015 of education partners who planned for a pilot of, what the Ministry branded, the Catch-Up (CU) program. The funds for the 80-school pilot were sourced by the Ministry from the Global Partnership for Education with additional support from UNICEF and J-PAL's Government Partnership Initiative (GPI). The Ministry also partnered with VVOB – education for development for implementation support in this pilot phase.

The pilot was implemented in 80 schools in 2016-2017. It confirmed that most grade 3 to 5 learners in Zambia were lacking basic reading and mathematics skills, with more than half of children in grades 3 to 5 in CU pilot schools at baseline unable to read words. The process evaluation results found that the program was well implemented, which is essential in establishing its feasibility. Critically, learning outcomes improved markedly during the one-year pilot period. For example, the share of children reading with basic proficiency (a simple paragraph or a story) grew by 18 percentage points from 34% to 52%.

The Ministry decided to expand the program after the pilot. The CU Program was awarded a grant by USAID Zambia to expand the program to 1800 schools (two provinces) over three years (2018-2020). Then again, in 2020, MoGE issued a letter on its plan to roll out CU to the remaining eight provinces. Following this, funding was secured for 8 of the 10 provinces from the LEGO Foundation, UNICEF, the Hempel Foundation, Co-Impact, and the Belgian government (DGD). As follows, four districts in Lusaka province received the CU program, starting implementation in early 2021. In 2022, the intervention was scaled up to all seven districts in Lusaka province. The three new districts in Lusaka province to Catch-Up are the focus of this study. Finally, in 2023, CU reached around 4,900 schools and approximately 735,000 children in grades 3-5 in eight provinces of Zambia.

1.2 The Program

CU is a remedial teaching program focusing on foundational numeracy and literacy among children enrolled in grades 3 to 5 of primary education. Key characteristics of the program are: (1) children are frequently assessed on literacy and numeracy skills using a standardized assessment tool. Using these assessments, children are grouped by learning level (and not by grade level). They can however easily transfer from one group to the next based on their performance on the assessment. (2) CU is a nonformal remedial teaching intervention whereas it takes place outside regular teaching hours. Often, the CU classes are organized over lunch or in the afternoon after the last class ended. And (3), the remedial teaching program is infused with elements of LtP. These elements of LtP are

positioned in a framework, in Zambia called the 7 Cs, which was originally based on the LEGO Foundation's characteristics of LtP. The 7Cs framework was developed by the MoGE in collaboration with the LEGO foundation to contextualize LtP in the Zambian culture and context. It recognizes 7 components of learning through play: concrete, captivating, connected, challenging, collaborative, creative, and cheerful. Concrete activities provide learners with the opportunity to actively engage with, manipulate and transform materials. Captivating activities attract and hold a learners' interest and allow the learner to make choices about the learning process. Connected activities relate to something already known by the learner, as well as being connected to their needs and interests. Challenging activities allow learners to discover things for themselves, instead of being provided by readymade solutions by the teacher. Collaborative activities allow learners to interact and collaborate with their peers. Creative activities give learners the possibilities to express their thoughts and emotions freely, without restrictions imposed by the teacher. Lastly, cheerful activities are activities that learners enjoy engaging in. LtP has been associated with socio-emotional development (Marbina et al, 2011; Zosh et al ,2017; Whitebread et al, 2017; Parker & Thomsen, 2019; Parker, van Beeck, & Callanan, 2019).

The main characteristics of CU as opposed to 'business as usual' teaching are described in Table 1.

Table 1: Comparison of 'Business as usual' to the Teaching at the Right Level methodology

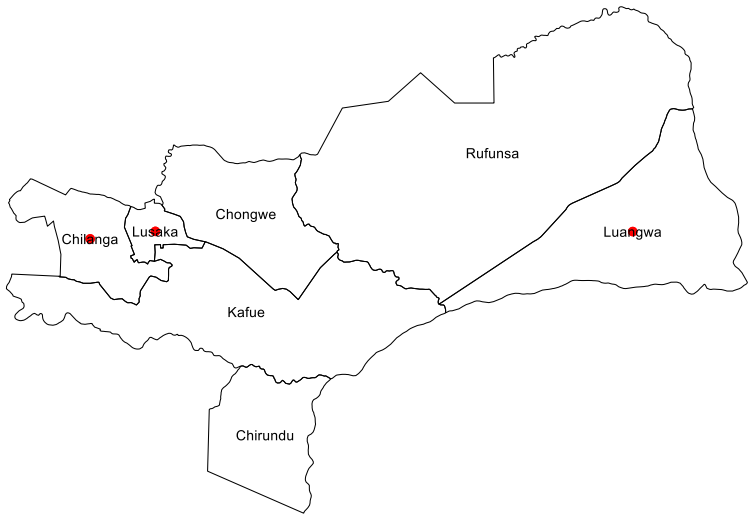
	'Business as usual'	Teaching at the Right Level (TaRL)
Goal & Assessment	Goal is to complete the grade level textbook or curriculum. Assessments aligned with curriculum and often complex – not suitable for classroom use.	Goal is to ensure basic foundational skills for all, with clearly articulated goals for basic reading and math. Simple one-on-one assessments of every student used for starting, grouping and tracking progress.
Training & Mentoring	Minimal continuous professional development or targeted coaching for teachers; traditionally non-practitioner trainers.	Practical training of leaders & instructors on the approach with practice periods of at least 20-25 days.
Grouping	Full class assembled together by grade level. Whole class teaching with little room for adjusting teaching to suit children's needs.	Children grouped by learning level rather than by grade. Children move quickly from one group to the next as their learning progresses.
Teaching & Learning Activities	Teachers focus mainly on whole-class instruction ('Chalk and Talk' or textbook-driven), focus on curriculum leads to teaching to the 'top of the class'.	Teachers use simple and engaging daily learning activities that can be adapted as children progress. Students engage in activities in large groups, small groups, and individually.
Measurement, Monitoring & Review	Pen and paper assessment done at the beginning and end of a learning unit. Minimal data analysis to understand student learning or adjust teaching before moving to the next learning unit.	Simple assessment to plan, and similar assessment used periodically to track student progress, review data, and make decisions on child progress and program design. Quick decision making to inform program delivery and future course.

2 | Methods

2.1 Participants and Context

All data presented here were collected within the scope of a quasi-experimental study, with a treatment and comparison group, on the impact of Catch-Up on socioemotional learning. To this end we collected data on children in grades 3 to 5 from three completely new districts in Lusaka province to the implementation of Catch-Up: Chilanga, Lusaka and Luangwa. Figure 2 presents the participating districts in the data collection on a map of Lusaka province.

Figure 2: Participating districts in rolling out the Catch-Up program



Because of full implementation in the districts across all schools at the same point in time, starting early 2022, we constructed a comparison group from the nearby districts in Central province. Chilanga and Lusaka districts are close to the border with Chibombo in Central province and have similar urban settings. The Luangwa district is a rural environment for which the Shibuyunji district in Central province may deliver comparable schools to the comparison group.

2.2 A priori Power Calculations

We performed a priori power calculations as to determine the number of schools to select to detect statistically significant effects on SEL of Catch-Up. Power calculations depend on several assumptions: the intra-class correlation coefficient (ICC); the number of students we can sample per cluster (school) and total number of clusters per arm; the predictive value of the model (R^2); voluntary or compulsory participation of teachers in Catch-Up (take-up); share of the students that we expect to spillover from treated to control schools; sample attrition; margin of error and confidence level.

Based on the previous literature, we position the ICC between 0.05 and 0.10 (Nickodem et al., 2019). Further, we sample about 30 learners per school (10 learners per grade), in the current scenario of 2 arms (i.e. a control and a treatment arm). Much less is known on the R^2 of the estimated models with SEL as the outcome variable: what exactly can predict good SEL-scores? The higher R^2 the

more we are assumed to predict the outcome SEL with a set of dependent variables (e.g. demographic characteristics). However, we observe only very small differences in the number of schools to sample whether R^2=0.35 or R^2=0.20.

Regarding take-up we argue it is equal to 1 (full take-up). As Catch-Up is implemented through the Ministry, teaching Catch-Up becomes part of the core duties of teachers – which makes it mandatory for the G3-5 teachers. We also argue that there will be no spillover of students from treated to control schools during the period of the research. Then again, sample attrition is fixed at 15%. As such, we assume that we shall not reach at endline about 15% of the students sampled at baseline. And finally, we use a traditional margin of error of 5% and a confidence level of 80% (Cohen, 1988).

Given these assumptions, and to reach a minimal detectable effect size (MDE) of 0.15 SD, we select 50 schools in the treatment group and 50 schools in the control group, having a sample size of approximately (N=3000) learners in both treatment and control schools together at baseline. Many interventions in education have a small impact in the short-term (Evans & Mendez-Acosta, 2021). This appears to hold true for SEL-related interventions in education; although it also heavily depends on the outcome indicators used for SEL (Taylor et al., 2017; Ganimian, 2020). It is then reasonable to assume $(MDE = 0.15\ SD)$ as to include enough schools in the study.

2.3 Design and Randomization

The 100 schools recruited in Lusaka (n=50) and Central province (n=50) were stratified by district, gender (5 boys and 5 girls per grade), and running agency (community schools, government schools, and grant-aided schools) and randomized at the school-level into the treatment and control group. According to the Zambia Education Management System (EMIS) data, for the control group, we have a total of 41 primary schools in Shibuyunji and 149 primary schools in Chibombo. For the treatment group we relied on the districts: Chilanga (107 schools); Luangwa (20 schools) and Lusaka (299 schools). As such, we randomly selected 13 schools in the Chilanga district, of which 8 schools are community schools, 4 government schools, and 1 grant-aided school. Regarding the Luangwa district, we randomly selected 2 government schools. Lusaka district is the largest district by far and we selected 35 schools at-random, of which 23 community schools, 11 government schools, and 1 grant-aided school. Regarding the control group, we randomly selected 31 schools from the 59 community schools in that district. Then again, 15 schools are run by the government and 2 schools are grant-aided. From the Shibuyunji district we randomly selected 2 government schools. Table 2 summarises the sample selection.

Table 2: Sample selection

Treatment Group (New districts to Catch-Up in Lusaka Province)					
Total sample			Sample by Running agency		
District	Schools	Sample of schools (Sample of students)	Community	GRZ	Grant-aided
Chilanga	114	13 (390)	8	4	1
Luangwa	19	2 (60)	0	2	0
Lusaka	318	35 (1050)	23	11	1
Total	**451**	**50 (1500)**	**31**	**17**	**2**
Control Group (Neighbouring districts in Central Province)					
Total sample			Sample by Running agency		
District	Schools	Sample of schools (Sample of students)	Community	GRZ	Grant-aided
Chibombo	149	48 (1440)	31	15	2
Shibuyunji	41	2 (60)		2	
Total	**190**	**50 (1500)**	**31**	**17**	**2**

Note: Schools from the rural district Shibuyunji should serve as a control group for schools from the Luangwa district. GRZ means government schools and grant-aided are government schools supported by grants (e.g. Church Organizations).

2.4 Assessment of Socioemotional Learning Domains

SEL encompasses competencies such as emotional regulation, perseverance, empathy, stress management, conflict resolution, self-concept, cooperation, ability to seeking help, self-regulation and leadership (D'Sa & Krupar,2021; Maguire, 2016; Pennequin, 2020; Taylor et al., 2017; Feinstein, 2015; Greenberg, 2010). Many questionnaires were developed by scientists to assess SEL domains (Taylor et al., 2017). We have chosen for the International Social Emotional Learning Assessment tool (ISELA), of which the psychometric properties are discussed by D'Sa & Krupar (2021). This questionnaire was chosen among other valuable questionnaires, because the tool has been shown to work by D'Sa & Krupar (2021) in low-resource contexts which is characteristic of most of the study sample. In addition, ISELA is a tool that is easy to contextualise and flexible to adapt; it can easily be used among respondents with low literacy levels; does not have restrictive copyright conditions; and allows us to assess a wide array of SEL domains (Barblett & Maloney, 2010).

Furthermore, the ISELA was developed for the age group we are targeting: young learners in grades 3 to 5. Many other validated SEL instruments are developed for older respondents. The ISELA is used to understand child development regarding self-concept, stress management, perseverance, empathy, and conflict resolution in children between 6 to 12 years. The definitions of these domains are (D'Sa & Krupar, 2021, p.28): self-awareness – the child's ability to express personal preferences, feelings and abilities; stress management – the conscious use of personal skills and resources to reduce the impact of stress; perseverance – the child's ability to stay on a task despite that task being difficult or delays in achieving success; empathy – an awareness and understanding of the emotions and expectations of others; relationships[1] – an understanding of the child's social networks and support

1 D'Sa & Krupar (2021, p.29): The domain 'relationship' is not measuring a SEL domain, however, answers to these questions are used to construct the other SEL-domains.

system; conflict resolutions – a child's interpersonal conflict resolution strategies. The scale measuring conflict resolutions has been divided into two subdomains: story – children's ranking of responses to stories of interpersonal conflict; and network – whether participants asked a family member, friend, or community member for help when resolving a peer issue.

ISELA further includes questions assessing the SEL learning environment of the child. Learning environment safety is defined as the safe and supportive environment around the child. The psychometric properties were not assessed by D'Sa & Krupar (2021), as it is not a SEL domain. We therefore consider an assessment of learning environment safety beyond scope of this study

2.5 Data Collection

The Centre for Promotion of Literacy in Sub-Saharan Africa (CAPOLSA, University of Zambia) organized the data collection in Lusaka and Central province, with logistical support of the international organization VVOB. There were two rounds of data collection. The first round started in February 2022 (baseline) and the second round in September 2022 (endline). Each round of data collection took about four weeks.

Both at baseline and endline CAPOLSA organized a SEL data collection workshop for assessors the week before the field work. There were 5 teams in total with 5 members in each team (1 supervisor and 4 assessors). Topics included at the workshops are: an understanding of the purpose and design of the Catch-Up program, an understanding of Child Protection and Research Ethics, principles of data collection, mastering ISELA and its use in the assessment of socioemotional learning among young children, and an understanding of play-based methods and approaches to learning. ISELA was taught and discussed in-depth, including role play and feedback, and assessment items were discussed and reviewed section by section in English, ciTonga, iciBemba and ciNyanja. The cultural appropriateness of each assessment item was also reviewed.

Prior to data collection, CAPOLSA visited the district education offices and presented the letters of authority (granting permission to conduct the study) from the Permanent Secretary and (stamped by the) Provincial Education Officer (PEO). All children participating in the study were asked for informed consent. Participation was voluntary and a child could always refuse collaboration during the survey.

Children were individually assessed by a trained assessor using tablets (KoboToolbox). The assessors conducted one-to-one verbal assessments, taking the child to a calm place in the neighborhood of the school (often the playground). The assessment took about 30 to 45 minutes per child. At the end of the day the assessors uploaded the data to safely store the data in a central, digital environment.

Demographic and socioeconomic characteristics of the child and his family were collected at baseline. Hereto CAPOLSA supplemented ISELA with questions relevant to the Zambian context. The questions were asked to the child. The demographic and socioeconomic characteristics include: gender, grade, a proxy for age[2], the language that the child feels comfortable speaking, whether the child has had breakfast before school, the employment status of the mother and whether the mother can read, the number of children in the household, or whether an adult from the family has been away from home for a long time. We also constructed an asset index which is a standardized index measuring the child's wealth based on (family) possessions. The asset index is based on the questions: do you have… a television, a stove in the house, electricity or solar power at home, a flushable toilet at home, a car, at least 2 sets of clothes, at least 1 set of shoes, a radio, a bed or mat to sleep on at home, a house with a cemented or tiled floor, a house with an iron roof or sheets, a cell phone, and a bicycle.

2 Exact birth dates are often difficult to assess by the children. In those cases the assessors estimated the child's approximate age.

Figure 3: Participant flowchart

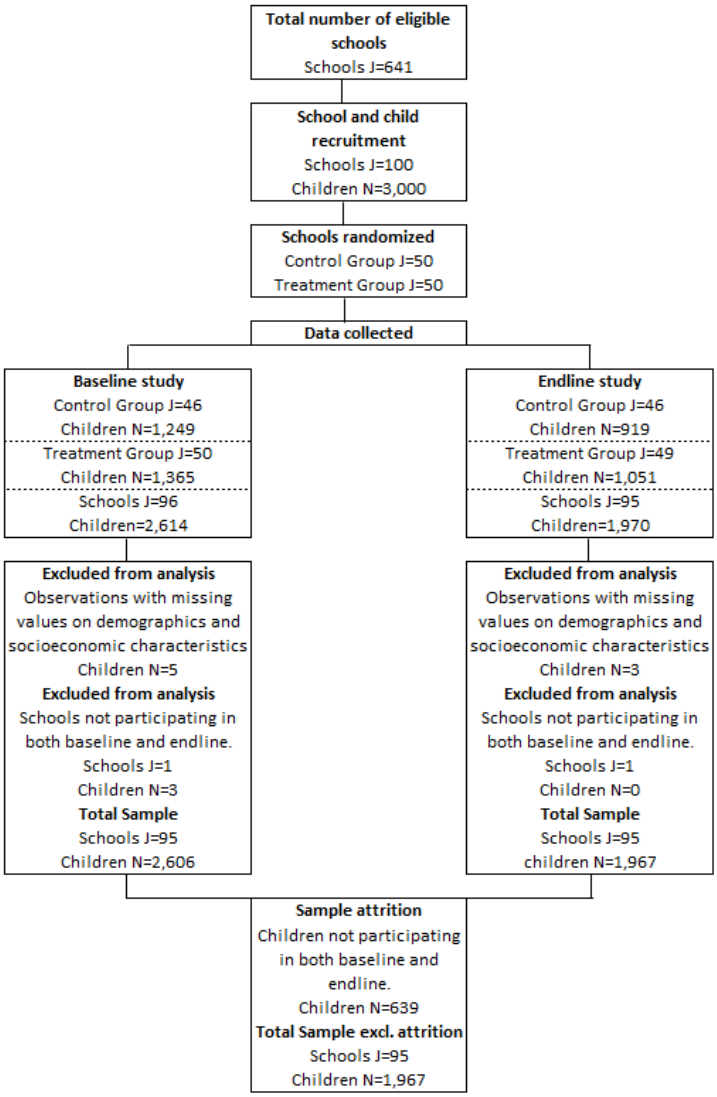


Figure 3 presents the data collection using a participant flow chart. While the initial target was to collect data from 3,000 children from 100 schools, the assessors could collect 2,614 child observations in the baseline study. However, because of data cleaning on missing information regarding demographic and socioeconomic characteristics, and missing information on the SEL-assessments, we lost 8 child observations. The final sample at baseline then consists of 2,606 child observations. This means a loss in sample size of 13,1% of the initial target of 3,000 children. Of these children the assessors have completed all SEL-assessments at the baseline.

In the endline study we could track 1,967 children back again. As such, compared to the baseline study, we could not collect SEL-assessments of 639 children, or one in every four children, in the endline study.

2.6 Sample Attrition and Missing Data

95 schools were retained in the endline study and included in the final sample of (*N*=2,606) child observations. Of the final sample, 639 children were assessed at baseline, but could not be traced back again at endline. The main reasons why are absenteeism on the day of assessment and moving

to a government school close by. Sample attrition may be a potential threat to the initial idea of random selection of schools, but only when children who dropout from the sample are very different from children observed in the full sample.

Table 3: Selectivity in sample attrition: descriptive statistics of the demographic and socioeconomic characteristics at baseline (*N*=2,606)

	No attrition (*N*=1,967)		Sample attrition (*N*=639)		Difference	Significance
	Mean	Std.Dev.	Mean	Std.Dev.		
Demographic characteristics						
Proxy age (in years)	10.6	1.7	10.9	1.8	0.347	***
Gender (1.Female)	0.506	0.500	0.501	0.500	-0.006	
Language						
Nyanja	0.588	0.492	0.581	0.494	-0.008	
Tonga	0.156	0.363	0.189	0.392	0.034	**
Other	0.256	0.437	0.230	0.421	-0.026	
Grade						
Grade3	0.334	0.472	0.354	0.478	0.020	
Grade4	0.324	0.468	0.349	0.477	0.025	
Grade5	0.342	0.474	0.297	0.457	-0.044	**
Household characteristics						
Breakfast	0.758	0.428	0.715	0.452	-0.043	**
Mother employed (1.No)	0.343	0.475	0.357	0.479	0.014	
Mother can read (1.Yes)	0.428	0.495	0.366	0.482	-0.061	***
Children in household	4.3	2.2	4.5	2.4	0.202	*
Adult away from home(1.Yes)	0.339	0.473	0.354	0.478	0.015	
Asset index	0.023	0.462	-0.070	0.503	-0.093	***

Significance denoted at the 10%-level (*), 5%-level (**) or 1%-level (***).

Table 3 summarizes the descriptive statistics on the statistical differences between the sample, assessed in both baseline and endline (*N*=1,967) and the sample only assessed at baseline, but not at endline (*N*=639). There are indeed a few variables statistically significantly different. Older children were more likely to drop out from the sample (the proxy in age, grade 5), as well as children with less favourable background characteristics (cf. mother cannot read or asset index). The estimated differences are limited in magnitude.

Table 4 further summarizes the statistical differences between treatment and control groups regarding the standardised SEL-domains assessed at baseline. We do not retain any significant difference between the treatment group and the control group.

Albeit limited in magnitude we flag a potential threat to the robustness of the effectiveness estimates to selective sample attrition and, as such, to missing data. We therefore suggest accounting for a missing data problem in the empirical strategy, among other things, by preserving the full sample for analysis and by offering a robust-correction approach to missing data in a randomized setting (Tubbs Dolan et al, 2022; Yeon Kim et al., 2023). We explain the empirical strategy as follows.

Table 4: Selectivity in sample attrition: descriptive statistics of the standardised outcome variables at baseline (*N*=2,606)

	No attrition (*N*=1,967)		Sample attrition (*N*=639)		Difference	Significance
	Mean	Std.Dev.	Mean	Std.Dev.		
Self concept	-0.011	1.015	0.010	1.019	0.021	
Stress management	-0.014	1.012	-0.061	1.016	-0.047	
Empathy	-0.211	0.993	-0.260	1.004	-0.049	
Perseverence	-0.036	1.077	-0.072	1.057	-0.036	
Story	-0.074	0.993	-0.157	0.980	-0.083	
Network	-0.011	1.013	0.005	1.015	0.016	
Conflict	-0.062	1.006	-0.117	1.013	-0.055	

Significance denoted at 5%-level (**).

2.7 Empirical Strategy

We wish to estimate the impact of the Catch-Up program on the assessed SEL-domains. To this end, we estimate a treatment effects model using the 'teffects ipwra' command in STATA. The command offers a robust-correction approach to missing data in a randomized setting. This missing data correction was also suggested in Tubbs Dolan et al. (2022).

In summary, two separate models are estimated: (1) a treatment model computing inversed-probability weights (ip); and (2) an outcome model performing a regression adjustment (ra). Regarding the first part of the command, the treatment model, the ip weights 'magnify' children from the treatment group who resemble children from the control group, and children from the control group who resemble children from the treatment group. The ip-weights are created estimating a probit model with treatment assignment as an outcome variable and baseline demographic and socioeconomic characteristics in the list of independent variables. The weights are thus based on propensity scores or the probability of treatment selection based on observed baseline characteristics (Austin & Stuart, 2015). As such, the ip weights facilitate the comparability of children in the treatment and control group in a similar vein as in propensity score matching models (Austin & Stuart, 2015). In fact, we will also estimate the weights using propensity score matching with an Epanechnikov Kernel specification as a robustness check to the treatment model. We find no differences in the conclusions made on balancing properties and overall treatment model performance as compared to standard ip weights of the 'teffects ipwra' command.

The regression adjustment (outcome) model implies the estimation of two separate regressions for the control and treatment groups. Each regression predicts treatment-specific outcomes at endline for each child, then computes the means of these predicted outcomes, and finally subtracts the means of the treatment group with the control group. The difference between the potential outcome means of the treatment group and the control group is the intention-to-treat effect (when full sample of N=2,606 is preserved) or the average treatment effect (dropping the 639 observations that do not have an endline score on the SEL domains) (Tubbs Dolan et al, 2022; Yeon Kim et al., 2023). The advantage of the 'ipwra' estimator is that it has double robustness properties: even if one of the two models is mis-specified, the treatment or outcome model, the estimator is still consistent.

We include at least the baseline SEL-assessments in the outcome model, taking account for potential (different) natural growth patterns in the treatment and control group between baseline and endline. Besides gender we also include age and/or grade in the outcome model and the socioeconomic characteristics to check for robustness of the results. There is a strong correlation ($\rho = 0.5$) between age and grade and we wish to avoid multicollinearity by including only one of those variables in the regression.

Finally, we transformed the scores on the SEL domains to a standardised scale with mean 0 and standard deviation of 1. We are then interested in the relative increases of children' SEL over time and between treatment and control groups. This approach further facilitates the comparability of effect sizes between the SEL domains.

3 | Results

3.1 Balancing properties

Table 5 presents the data collected on the full sample of (*N*=2,606) child observations together with the statistical differences between the treatment and control groups. The variables included in this table shall also be used as control variables in the analysis.

We use information on the child's gender and grade to test whether stratification in data collection happened successfully. We confirm that there are no statistical differences between the treatment and the control group on those two variables.

Table 5: Demographic and socioeconomic characteristics of the sample (N=2,606)

	Control Group (*N*=1,249)		Treatment Group (*N*=1,357)		Diff.	Significance
	Mean	Std.Dev.	Mean	Std.Dev.		
Gender (1.Female)	0.498	0.500	0.511	0.500	0.013	
Grade						
Grade3	0.337	0.473	0.340	0.474	0.003	
Grade4	0.336	0.473	0.325	0.469	-0.011	
Grade5	0.327	0.469	0.335	0.472	0.008	
Breakfast (1.Yes)	0.751	0.433	0.744	0.436	-0.007	
Mother employed (1.No)	0.339	0.474	0.352	0.478	0.013	
Mother can read (1.Yes)	0.376	0.484	0.447	0.497	0.071	***
Children in household	4.685	2.526	3.991	1.936	-0.693	***
Adult away from home (1.Yes)	0.344	0.475	0.340	0.474	-0.004	
Asset index	-0.237	0.456	0.218	0.375	0.455	***

Note. Significance denoted at the 10%-level (*), 5%-level (**) or 1%-level (***).

Despite the random selection of schools, and successful stratification, we observe significant differences regarding the socioeconomic characteristics. Children from the treatment group have significant more mothers who can read than children in the control group. They live with fewer children in the household and have a more favourable asset index.

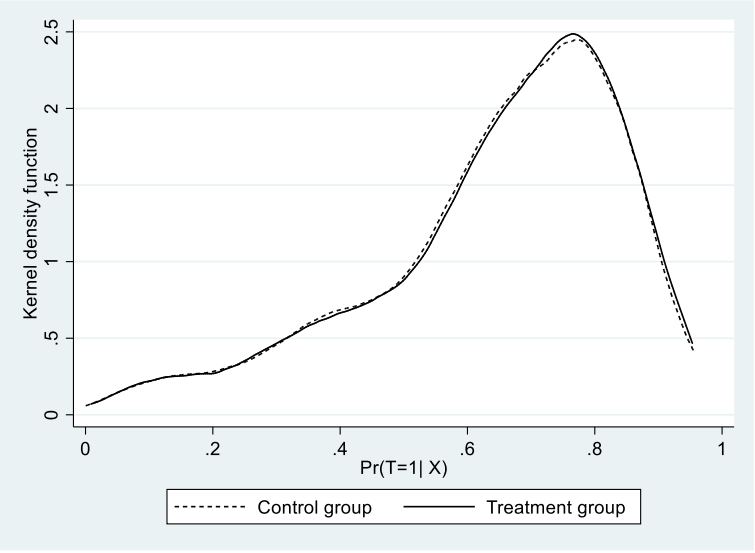
Table 6 then again presents the robust-correction approach to missing data using ip weights. To this end, we estimated a probit model with treatment assignment as the outcome variable and baseline demographic and socioeconomic characteristics in the list of independent variables. Thanks to this method there are no longer significant differences observed on the variables in Table 6. Figure 4 also visually indicates that the bundle of covariates is overlapping across treatment and control groups, which gives us confidence in the empirical strategy chosen.

Table 6: Inversed-probability weighting applied to the demographic and socioeconomic characteristics of the sample (*N*=2,606)

	Control Group (*N*=1,249)		Treatment Group (*N*=1,357)		Difference	Significance
	Mean	Std.Dev.	Mean	Std.Dev.		
Demographic characteristics						
Gender (1.Female)	0.511	0.500	0.511	0.500	0.001	
Grade					0.000	
Grade3	0.366	0.482	0.340	0.474	-0.025	
Grade4	0.325	0.468	0.324982	0.468541	0.000	
Grade5	0.310	0.462	0.335	0.472	0.025	
Household characteristics						
Breakfast (1.Yes)	0.743	0.437	0.744	0.436	0.002	
Mother employed (1.No)	0.343	0.475	0.352	0.478	0.009	
Mother can read (1.Yes)	0.461	0.499	0.447	0.497	-0.014	
Children in household	3.997	2.011	3.991	1.936	-0.005	
Adult away from home (1.Yes)	0.369	0.483	0.340	0.474	-0.028	
Asset index	0.214	0.376	0.218	0.375	0.005	

Note. Significance denoted at the 10%-level (*), 5%-level (**) or 1%-level (***).

Figure 4: Balancing properties of the bundle of covariates across treatment and control groups



Note. All variables listed in Table 6 are included in the bundle of covariates.

3.2 Internal consistency

We check the internal consistency reliability using measures of Cronbach's alpha. This statistic indicates to what degree questions in a scale measure the same underlying concept consistently (e.g., whether respondents responded to the questions in a consistent manner). As a rule of thumb, the Cronbach's alpha should be equal to or above 0.7 to conclude reliability. This holds true for all of the five SEL-domains.

Table 7: Internal consistency reliability

	Baseline study	Endline study
Self-concept	0.92	0.93
Stress management	0.76	0.87
Empathy	0.70	0.91
Perseverance	0.77	0.83
Conflict resolution	0.73	0.84
Story	0.73	0.81
Network	0.72	0.81

3.3 Change in Socioemotional domains

Table 8 presents a summary of the main effects (in SD) of the Catch-Up program on the SEL-domains. We estimated three models in total with each model increasing in the number of control variables in the outcome model. In Model 1 we only include the baseline SEL-assessment, in Model 2 we add the variables grade and gender (on which we stratified the data collection), and in Model 3 we add a set of socioeconomic characteristics (see Table 5). All models estimate robust standard errors clustered at the school-level (i.e. 95 clusters in total). At the bottom of Table 8 we also present the intraclass correlation coefficient (ICC). The ICC is in all models estimated close to 0, meaning that the SEL-outcomes of children are quite individualistic achievements. There is as such no evidence as such that multilevel models would perform better than the empirical strategy chosen in this paper.

Across the three models we find strong robust evidence that the Catch-Up program positively influenced the empathy domain. We estimate an effect of 0.15 SD significant at 1%-level. We also find a positive impact on the sub-scale 'story' of the conflict resolution domain. The main effect is equal to 0.11 SD significant at 5%-level. At the same time, we find a negative impact of -0.08 SD significant at 5%-level on the sub-scale 'network' of the conflict resolution domain.

We perform several robustness checks on these results, including dropping the predictions of the endline results of children participating only in the baseline study (N=639) and estimating matching models without regression adjustments. Results and significance-level on the empathy and story domain hold, while the network domain remains significant at 10%-level.

We also perform an analysis on the heterogeneity of the effects across boys and girls, school types (community or government schools) and poverty status (above or below the median asset index). These results are presented in Table 9. We find that the Catch-Up program had the largest effects on children in community schools. SEL-domains empathy and story regarding the group of male children, and children with a wealth status above the median, were also more impacted. The negative impact on network appears again among girls and children with a wealth status below the median.

In summary, expressing the main effects in effect sizes using Cohen's d, we find effect size of $d = 0.141\ SD$ on the empathy domain, $d = 0.121\ SD$ on the story domain and $d = -0.083\ SD$ on the network domain. Catch-Up did not have had any significant impact on the other three domains self-concept, stress management and perseverance.

Table 8: Summary of the changes in socioemotional domains (*N*=2,606): Main Effects

	Self concept	Stress management	Empathy	Perseverance	Story	Network	Conflict
Model 1							
Impact (1.Treat)	0.095	0.000	0.153 ***	0.011	0.112 **	-0.087 **	0.041
	(0.062)	(0.052)	(0.056)	(0.045)	(0.065)	(0.050)	(0.064)
Control variables	Assess T-1	Assess T-1	Assess T-1	Assess T-1	Assess T-1	Assess T-1	Assess T-1
Outcome model							
Model 2							
Impact (1.Treat)	0.077	0.002	0.147 ***	0.005	0.113 *	-0.091 *	0.039
	(0.054)	(0.051)	(0.054)	(0.040)	(0.060)	(0.049)	(0.060)
Control variables	Assess T-1	Assess T-1	Assess T-1	Assess T-1	Assess T-1	Assess T-1	Assess T-1
Outcome model	Gender	Gender	Gender	Gender	Gender	Gender	Gender
	Grade	Grade	Grade	Grade	Grade	Grade	Grade
Model 3							
Impact (1.Treat)	0.037	0.011	0.144 ***	-0.004	0.123 **	-0.085 *	0.048
	(0.045)	(0.050)	(0.051)	(0.029)	(0.056)	(0.048)	(0.056)
Control variables	Assess T-1	Assess T-1	Assess T-1	Assess T-1	Assess T-1	Assess T-1	Assess T-1
Outcome model	Gender	Gender	Gender	Gender	Gender	Gender	Gender
	Grade	Grade	Grade	Grade	Grade	Grade	Grade
	Breakfast	Breakfast	Breakfast	Breakfast	Breakfast	Breakfast	Breakfast
	Employment	Employment	Employment	Employment	Employment	Employment	Employment
	Reading	Reading	Reading	Reading	Reading	Reading	Reading
	Children HH	Children HH	Children HH	Children HH	Children HH	Children HH	Children HH
	Adult away	Adult away	Adult away	Adult away	Adult away	Adult away	Adult away
	Asset index	Asset index	Asset index	Asset index	Asset index	Asset index	Asset index
Standard error	Clustered	Clustered	Clustered	Clustered	Clustered	Clustered	Clustered
ICC	0.023	0.013	0.047	0.017	0.036	0.013	0.034

Note: Significance denoted at the 10%-level (*), 5%-level (**) or 1%-level (***).

Table 9: Summary of changes in socioemotional domains (*N*=2,606): Heterogeneity of the effects

	Self concept		Stress management		Empathy			Perseverance	
	0.Community	1.Community	0.Community	1.Community	0.Community	1.Community		0.Community	1.Community
Impact (1.Treat)	-0.010	0.066	-0.025	0.060	0.015	0.254	***	-0.014	-0.028
	(0.079)	(0.057)	(0.064)	(0.082)	(0.079)	(0.059)		(0.044)	(0.039)
	0.Boy	1.Girl	0.Boy	1.Girl	0.Boy	1.Girl		0.Boy	1.Girl
Impact (1.Treat)	-0.001	0.068	0.029	0.013	0.184	0.114	**	0.005	-0.003
	(0.054)	(0.064)	(0.069)	(0.059)	(0.058)	(0.064)		(0.042)	(0.034)
	0.poor	1.poor	0.poor	1.poor	0.poor	1.poor		0.poor	1.poor
Impact (1.Treat)	-0.072	0.067	0.007	-0.002	0.178	0.122		0.020	-0.043
	(0.059)	(0.063)	(0.064)	(0.049)	(0.049)	(0.083)		(0.038)	(0.037)

	Story		Network		Conflict		
	0.Community	1.Community	0.Community	1.Community	0.Community	1.Community	
Impact (1.Treat)	0.035	0.181	-0.134	-0.026	-0.043	0.117	*
	(0.089)	(0.061)	(0.066)	(0.075)	(0.087)	(0.067)	
	0.Boy	1.Girl	0.Boy	1.Girl	0.Boy	1.Girl	
Impact (1.Treat)	0.161	0.076	-0.066	-0.097	0.088	0.008	
	(0.073)	(0.069)	(0.066)	(0.058)	(0.076)	(0.067)	
	0.poor	1.poor	0.poor	1.poor	0.poor	1.poor	
Impact (1.Treat)	0.158	0.138	-0.067	-0.111	0.087	0.041	
	(0.065)	(0.069)	(0.064)	(0.055)	(0.070)	(0.059)	

Note: All models cluster the standard error at the school-level. All models control for demographic and socioeconomic characteristics.
Significance denoted at the 10%-level (*), 5%-level (**) or 1%-level (***).

4 | Conclusion and discussion

The findings show that the Catch-Up program enhanced children's capacity to understand and empathize with others, as well as their ability to navigate and resolve conflicts in social settings. The magnitudes of the significant effect sizes are small, however, but in line with what to expect from educational interventions that at time of evaluation only ran for 6-7 months (Merchie et al., 2018; Popova et al., 2022). These findings may arise from a strong focus on group work in CU (see Deliverable II). Collaborating in groups will inherently lead to exposure to conflict situations, which gives children the opportunity to navigate conflict in the relatively safe setting of a classroom setting, with the guidance of a teacher. No significant impacts were found on the domains of self-concept, stress management, and perseverance. This implies that while Catch-Up may have shown some encouraging improvements in three SEL-domains, further research is needed to establish a conclusive relationship between remedial teaching infused with learning through play and domains of socioemotional learning.

Learning through play and SEL are common concepts in early childhood education but are still uncommon practices in primary education in Zambia. Deliberate programming from the Zambian Ministry of Education to promote Social Emotional Learning at primary school levels is missing. The Catch-Up remedial program provides an opportunity for learners in primary education to benefit from its activities with their characteristics of learning through play.

A few limitations of this study are worth mentioning. First, there was only a relatively short exposure time to the Catch-Up program of 7-9 months, at least, that we could evaluate, because implementation also ran beyond this evaluation study. The research team could not (re-)evaluate at later points in time because of expected implementation of the Catch-Up program in Central province, which of course would impact the control group. Despite the short duration of 7-9 months, the results indicated small but significant impacts on a few SEL domains. Nonetheless, literature has shown that sustained exposure to SEL interventions leads to more gains in social and emotional skills of children (e.g., Nelson et al., 2003). For example, Hunter, DiPerna, Cheng, Lei, & Hart (2021) showed that compared to students who experienced an SEL intervention in second grade only, students exposed to this intervention in both first and second grade showed further gains in their social and emotional skills. In order to be effective, several meta-analyses suggest that programs should be of a certain length or duration, somewhere between 3-6 months (with weekly classes). However, this duration might be insufficient to obtain long-term effects if no later booster sessions are held (see Kraag et al., 2006, Weissberg et al., 2007). It is noteworthy that the meta-analysis of Kraag et al. (2006) specifically focused on stress management, which is one dimension that failed to reach significant effects in this study.

References

Aber, J. L., Torrente, C., Starkey, L., Johnston, B., Seidman, E., Halpin, P., ... & Wolf, S. (2017). Impacts after one year of "Healing Classroom" on children's reading and math skills in DRC: Results from a cluster randomized trial. *Journal of Research on Educational Effectiveness*, *10*(3), 507-529.

Asio, J. M. R., & Jimenez, E. C. (2020). Effect of Remediation Activities on Grade 5 Pupils' Academic Performance in Technology and Livelihood Education (TLE). Pedagogical Research, 5(4), em0075. https://doi.org/10.29333/pr/8464

Austin, P. C., & Stuart, E. A. (2015). Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Statistics in medicine*, *34*(28), 3661-3679.

Banerjee, A. V., Cole, S., Duflo, E., & Linden, L. (2007). Remedying education: Evidence from two randomized experiments in India. *The quarterly journal of economics*, *122*(3), 1235-1264.

Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukherji, S., ... & Walton, M. (2016). *Mainstreaming an effective intervention: Evidence from randomized evaluations of "Teaching at the Right Level" in India* (No. w22746). National Bureau of Economic Research.

Barblett, L., & Maloney, C. (2010). Complexities of assessing social and emotional competence and wellbeing in young children. *Australasian Journal of Early Childhood*, *35*(2), 13-18.

Broadhead, P. (2006). Developing an understanding of young children's learning through play: the place of observation, interaction and reflection. *British Educational Research Journal*, *32*(2), 191-207.

Brown, L. E., Kim, H. Y., Tubbs Dolan, C., Brown, A., Sklar, J., & Aber, J. L. (2023). Remedial programming and skill-targeted SEL in Low-Income and crisis-affected contexts: Experimental evidence From Niger. *Journal of Research on Educational Effectiveness*, 1-32.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates

D'Sa, N., & Krupar, A. (2021). Developing and Validating the International Social and Emotional Learning Assessment: Evidence from a Pilot Test with Syrian Refugee Children in Iraq. *Journal on Education in Emergencies* 7 (2), DOI: 20-56. https://doi.org/10.33682/xdpq-bwp2.

Evans, D. K., & Mendez Acosta, A. (2021). Education in Africa: What are we learning?. *Journal of African Economies*, *30*(1), 13-54.

Hamre, B. K., & Pianta, R. C. (2005). Can instructional and emotional support in the first-grade classroom make a difference for children at risk of school failure?. *Child development*, *76*(5), 949-967.

Hunter, L. J., DiPerna, J. C., Cheng, W., Lei, P., & Hart, S. C. (2021). Twice as Nice? Sustained Exposure to a Universal Social–Emotional Learning Program Across Multiple Grades. *School Mental Health*, *13*(1), 84-100.

Kim, H. Y., Brown, L., Tubbs Dolan, C., Gjicali, K., Deitz, R., Prieto Bayona, M. D. S., & Aber, J. L. (2023). Testing the impact of a skill-targeted social and emotional learning curriculum and its variation by pre-and postmigration conflict experiences: A cluster randomized trial with Syrian refugee children in Lebanon. *Journal of Educational Psychology*.

Kraag, G., Zeegers, M. P., Kok, G., Hosman, C., & Abu-Saad, H. H. (2006). School programs targeting stress management in children and adolescents: A meta-analysis. *Journal of school psychology*, *44*(6), 449-472.

Lakshminarayana, R., Eble, A., Bhakta, P., Frost, C., Boone, P., Elbourne, D., & Mann, V. (2013). The Support to Rural India's Public Education System (STRIPES) trial: A cluster randomised controlled trial of supplementary teaching, learning material and material support. *PloS one*, *8*(7), e65775.

Lungu, S., & Matafwali, B. (2020). Play based learning in early childhood education (ECE) centres in Zambia: A teacher perspective. *European Journal of Education Studies*, *7*(12).

Marbina, L., Chur, A., Tayler, C., & Church, A (2011). Practice Principle 6: Integrated teaching and development Framework.

Merchie, E., Tuytens, M., Devos, G., & Vanderlinde, R. (2018). Evaluating teachers' professional development initiatives: towards an extended evaluative framework. *Research papers in education*, *33*(2), 143-168.

Metsäpelto, R. L., & Pulkkinen, L. (2014). The benefits of extracurricular activities for socioemotional behavior and school achievement in middle childhood: An overview of the research. *Journal for educational research online*, *6*(3), 10-33.

Moran, P. R. (1971). The integrated day. *Educational Research*, *14*(1), 65-69.

Nelson, G., Westhues, A., & MacLeod, J. (2003). A meta-analysis of longitudinal research on preschool prevention programs for children. *Prevention & Treatment*, *6*(1), 31a.
Panayiotou, M., Humphrey, N., & Wigelsworth, M. (2019). An empirical basis for linking social and emotional learning to academic performance. *Contemporary Educational Psychology*, *56*, 193-204.
Parker, R., van Beeck, L., & Callanan, J. (2019). Measures of characteristics and skills associated with learning through play: Findings from the literature.
Parker, R., & Thomsen, B. S. (2019). Learning through play at school: A study of playful integrated pedagogies that foster children's holistic skills development in the primary school classroom.
Popova, A., Evans, D. K., Breeding, M. E., & Arancibia, V. (2022). Teacher professional development around the world: The gap between evidence and practice. *The World Bank Research Observer*, *37*(1), 107-136.
Puerta, M. L. S., Valerio, A., & Bernal, M. G. (2016). Taking stock of programs to develop socioemotional skills: A systematic review of program evidence. Directions in Development, Washington DC: World Bank. Doi: 10.1596/798-1-4648-0872-2.
Selvarajan, P. (2022). The impact of remedial teaching on improving the competencies of low achievers. *International Journal of Social Science & Interdisciplinary Research* 2277-3630, 11(01), 283-287.
Taylor, R. D., Oberle, E., Durlak, J. A., & Weissberg, R. P. (2017). Promoting positive youth development through school-based social and emotional learning interventions: A meta-analysis of follow-up effects. *Child development*, *88*(4), 1156-1171.
Torrente, C., Aber, J. L., Starkey, L., Johnston, B., Shivshanker, A., Weisenhorn, N., ... & Tubbs Dolan, C. (2019). Improving primary education in the Democratic Republic of the Congo: End-line results of a cluster-randomized wait-list controlled trial of Learning in a Healing Classroom. *Journal of Research on Educational Effectiveness*, *12*(3), 413-447.
Tubbs Dolan, C., Kim, H. Y., Brown, L., Gjicali, K., Borsani, S., Houchaimi, S. E., & Aber, J. L. (2022). Supporting Syrian Refugee children's academic and social-emotional learning in national education systems: A cluster randomized controlled trial of nonformal remedial support and mindfulness programs in Lebanon. *American Educational Research Journal*, *59*(3), 419-460.
Vromant, N., Kuppens, L., Hazemba, M., Mwamba, F. K., & Cupito, E. (2021). Scaling Teaching at the Right Level. *Ensuring All Children Learn: Lessons from the South on What Works in Equity and Inclusion*, 101.
Walker, L. (2021). Explorations of a Middle School Remedial Program Emphasizing Social and Emotional Learning (Doctoral dissertation, Brenau University).
Weissberg, R.P., Durlak, J.A., Taylor, R.D., Dymnicki, A.B., O'Brien, M.U. (2007). Promoting Social and Emotional Learning Enhances School Success: Implications of a Meta-analysis. Draft paper, May 28th, pp. 42.
Whitebread, D., Neale, D., Jensen, H., Liu, C., Solis, S. L., Hopkins, E., & Zosh, J. (2017). *The role of play in children's development: a review of the evidence*. White Paper, Billund, Denmark: LEGO Foundation.
Wolf, S., Torrente, C., Frisoli, P., Weisenhorn, N., Shivshanker, A., Annan, J., & Aber, J. L. (2015). Preliminary impacts of the "Learning to Read in a Healing Classroom" intervention on teacher well-being in the Democratic Republic of the Congo. *Teaching and Teacher Education*, *52*, 24-36.
Zosh, J. N., Hopkins, E. J., Jensen, H., Liu, C., Neale, D., Hirsh-Pasek, K., & Whitebread, D. (2017). *Learning through play: a review of the evidence*. White Paper, Billund, Denmark: LEGO Foundation.

Deliverable II: Implementation Fidelity and Teacher Surveys

1 | Methods

1.1 Participants and Context

Data was collected as part of a larger scale evaluation of the implementation of the Catch-Up program in Lusaka province in Zambia. The evaluation aimed to determine both learners' social and emotional learning (Deliverable 1), as well as the impact of the program on teacher's knowledge, attitudes and practices (current deliverable) and spill-over of Catch-Up promoted classroom practices in regular classes (Deliverable 3). The current study included teachers from both treatment and control schools to participate in a self-reported survey on their knowledge, attitudes, and practices concerning the use of Catch-Up promoted classroom practices in regular classes. For treatment schools only, enumerators also observed the implementation fidelity of the program in a Catch-Up class, to determine whether Catch-Up was implemented in the first place as well as the quality of implementation.

1.2 Design and Randomization

Data was collected in 3 districts in Lusaka province and 2 districts in Central province.[3] The Catch-Up program was implemented in all schools in Lusaka province, hence only treatment schools were sampled from Lusaka. Catch-Up was not yet implemented in the Central province at the time of the study, so all schools selected from there served as control schools. The total sample existed of 89 teachers, with 45 teachers from control schools and 44 teachers from treatment schools. Schools were selected from districts in Lusaka province that were new to Catch Up, and one teacher was sampled per school. Teachers who participated in the teacher surveys were the same teachers who participated in the endline classroom observations from Deliverable 3.[4] In addition to teacher surveys, enumerators also conducted observations of Catch-Up classes at treatment schools using the implementation fidelity tool.

1.3 Assessment of Constructs

Implementation fidelity tool. The implementation fidelity tool was completed by enumerators who observed Catch-Up classes at endline. The tool existed of 10 questions measuring whether Catch-Up recommended practices were implemented. Questions were developed by CAPOLSA, in collaboration with VVOB, based on Catch-Up guidelines that teachers are trained on and are supposed to adhere in their Catch-Up classes. Two questions were measured on a "Yes/No" scale, were the enumerator reported on the presence or absence of a specific classroom practice (example question *"The teacher gave learners individual activities"*). Five questions were measured on a 5-point Likert scale running from "Not at all" to "Very much" (example question *"The teacher began the class with a*

3 The districts were Chilanga, Luangwa, Lusaka, Chibombo, and Shibuyunji.

4 For classroom observations, it is important to note that sampling happened at the class level and not the teacher level. Catch-Up program implementation happens at the class level, meaning that if a Catch-Up class has a change in teacher, the new teacher assigned to that class is expected to take over Catch-Up implementation. For sampling this meant that in these situations, the new teacher would be sampled to be included in the endline. All teachers sampled in both Deliverable 2 and 3 are consequently expected to have participated in Catch-Up training. However, this might not have been the case for teachers who were only recently assigned to Catch-Up classes at the time of endline data collection.

whole class activity"). These five questions showed high internal consistency (Cronbach's a = .7916). The remaining 3 questions were measured on different scales and asked about the number of activities performed in class, the time the class takes place, and the duration of the class. The implementation fidelity tool can be found in the appendix.

Teacher survey. This self-reported survey existed of 4 sections, named Section A to Section D. Questions developed were based on several existing scales (Ajzen, 2020; Gokce, 2010; OECD, 2018), the 7C's framework of learning through play, and Catch-Up program documentation (training materials, teacher guides, etc.). During enumerator training it became clear from observing enumerators engaging with the survey that they understood the questions included in the survey, hence the research team did not deem pilot testing of the tool necessary.

Section A aimed to measure teachers' knowledge about the Catch-Up teaching process. Questions were phrased like statements, which all referred to Catch-Up promoted classroom practices. Some statements referred to best practices and ideas, while others referred to less desirable practices and ideas. The section existed of 11 questions, which were measured on a 4-point Likert scale running from "Strongly disagree" to "Strongly agree". Questions covered statements about the role of the teacher, learner problem solving strategies, who should control classroom activities, how teaching and the classroom should be structured, and how learners learn. Items were conceptually different, to capture a wide array of the different Catch-Up promoted classroom practices. Consequently, they did not have a high internal consistency (Cronbach's a = .5421). Example items are *"Learners should be allowed to think of solutions to practical problems themselves before the teacher shows them how they are solved"*, *"My role as a teacher is to facilitate the learner's own inquiry"* and *"A quiet classroom is generally needed for effective learning"*.

Section B aimed to measure teacher's self-reported implementation of Catch-Up promoted classroom practices, notably learning through play. Questions in this section aimed to cover all 7 components of the 7 C's framework (concrete, captivating, connected, challenging, collaborative, creative, cheerful), with some questions mapping on multiple components. Questions were measured on a 4-point Likert scale running from "Not at all" to "To a great extent". The questions showed high internal consistency (Cronbach's a = .8244). Example items are *"The class activities include songs and games"*, *"Learners are allowed to share their feelings and emotions freely in class"*, and *"Learners receive the opportunity to try again if they failed"*.

Section C measured teachers' job satisfaction. This section contained 7 questions, which were all measured on a 4-point Likert scale running from "Strongly disagree" to "Strongly agree". The questions within this section showed sufficient internal consistency (Cronbach's a = .6582). While not a direct target of the Catch-Up implementation, there is anecdotal evidence from program monitoring and ongoing research projects that teachers enjoy participating in the Catch-Up program. It would be interesting to establish if this translates into higher job satisfaction for the teacher job. On the other hand, job satisfaction could also reduce because of the additional work Catch-Up requests from teachers. As mentioned before, classes are supposed to be taught outside of regular class hours and require planning. Example items of these section are *"All in all, I am satisfied with my job"*, *"My job pays me well"*, and *"I feel that I am making a significant education difference in the lives of my learners."*

The questions in Section D were based on the Theory of Planned Behaviour (TPD) (Ajzen, 1991). TPD was added as a theorical framework because it helps to explain and predict human behaviour. Section D aimed to measure teachers' motivation and intention to implement Catch-Up, by asking questions about their personal attitudes, subjective norms, perceived behavioural control, behavioural intentions, and actual behaviour. It was measured by 8 items, of which 7 were retained in the eventual

analysis[5]. Items were all measured on a 5-point Likert scale running from "Strongly disagree" tot "Strongly agree". They had high internal consistency (Cronbach's a = .8148). Example items are *"I intend to continue teaching Catch-Up classes for at least the next 1 year"*, *"My school authorities do not approve of my participation in the Catch-Up program"*, and *"I feel that I do not have the necessary resources to teach on the Catch-Up program"*. The teacher survey is included in Annex 4.

1.4 Data Collection

Approval to conduct the study was obtained from the Ministry of Education, and ethical clearance was obtained from the Research Ethics Committee of the University of Zambia. Data was collected in September-November 2022 by the Centre for the Promotion of Literacy in Sub-Saharan Africa (CAPOLSA). Data in this study was collected during the endline phase of the research, which was around 7-9 months after Catch-Up was rolled out in Lusaka province. The implementation fidelity tool was filled out by trained enumerators themselves, as they observed a Catch-Up class. The teacher survey tool was administered by enumerators to teachers, and thus collected teachers' own responses. All data was collected on the Kobo collect app. All respondents participating in the study were asked for informed consent. Participation was voluntary and a respondent could always refuse collaboration during the survey. Data was downloaded in Excel format and converted to Stata datasets for data cleaning and analysis.

5 Question D8 *"I have not yet started teaching Catch-Up classes"* was removed from analysis, because it should have been measured on a "Yes/No" scale, instead of on a 5-point Likert scale running from "Strongly disagree" to "Strongly agree".

2 | Empirical Strategy and Results

The implementation fidelity and teacher surveys were analysed using Stata 18.

Implementation fidelity. At the beginning of implementation fidelity tool, teachers were asked if Catch-Up was being implemented in their school. Surprisingly, 10 of the 40 observations had a "Not Applicable" score for this question, while 3 observations stated "No". Unfortunately, on days that data collection teams visited the schools, Catch-Up teachers were not present or Catch-Up classes were not being implemented at some schools. Since the implementation fidelity tool was to be conducted during Catch-Up classes specifically, observations conducted at these schools could not be used. This led to a total sample size of 25 observations for the implementation fidelity tool.

Two questions were measured on a "Yes/No" scale, for which enumerators reported whether specific practices were implemented or not. The table below shows these questions and their corresponding results. At most schools these practices were reported to be implemented.

Question	**Yes**	**No**
Learners are grouped by learning level rather than by grade.	19 (76.0%)	6 (24.0%)
Teacher gave learners individual activities.	18 (72.0%)	7 (28.0%)

Next follows a group of 5 questions measured on a 5-point Likert scale running from "Not at all" to "Very much". While the first 4 items were observed and reported by the enumerator, the last item about mentorship/support from the school was asked directly to the teacher. The table below summarizes the results for these questions. Interestingly, the average scores on 4 out of 5 of these items are above 3, so they are at the high end of the scale. This shows that enumerators observed high implementation fidelity in the selected schools. For each item, a *t*-test was computed to compare the average score with the midpoint of the scale (a score of 3), to determine if the score was significantly different from the midpoint. Scores that are significantly different from the midpoint of 3 are further indication that teacher scores show high implementation fidelity, assuming that the data are normally distributed in the population. For 3 out of the 4 items with average scores above 3, significance was established.

Question	**M**	**SD**	***p*-value *t*-test**	**CI *t*-test**
The teacher ensures that activities capture the 7 Cs.	4.04	.6758	.000	3.761-4.319
The teacher began the class with a whole class activity.	4.36	.9074	.000	3.985-4.735
The teacher grouped learners in small groups for an activity.	3.24	1.739	.248	2.522-3.958
The teacher allows the group leader to lead some classroom activities.	2.72	1.838	.773	1.961-3.479
The teacher receives mentorship/support from school mentors.	3.80	1.190	.001	3.309-4.291

The tool had an additional 3 questions which were all measured on different scales. For one question, enumerators reported how many activities they observed during class. None of them reported the absence of activities, and 80% observed 2 or more activities during the class. This is a positive result, as Catch-Up classes are supposed to actively engaged learners.

Response options	Percentage of responses
No activity	0 (0%)
1 Activity	5 (20.0%)
2 Activities	9 (36.0%)
3 Activities	10 (40.0%)
More than 3 activities	1 (4%)

Catch-Up classes are supposed to be implemented outside of regular class hours. Teachers and schools are free to choose whether they want to schedule these classes before or after regular class hours. In Zambia, many schools schedule Catch-Up classes over lunch time or in the afternoon (after class ending). Not all students are a whole day at school because of a double shifting system implemented by many schools, with schooling being organised at different times of the day for different grades. Enumerators reported when the Catch-Up class they observed took place. As desired, most classes took place outside of regular class hours.

Response options	Percentage of responses
Before regular class	9 (36.0%)
During regular class	3 (12.0%)
After regular class	13 (52.0%)

Lastly, enumerators were asked to report how long the Catch-Up class took. It is recommended that these classes should last around 1 hour. The majority of classes did indeed take roughly an hour, but the finding of 36% of classes taking less or more time than the desired one hour warrants further investigation.

Response options	Percentage of responses
< 1 hour	6 (24.0%)
Approximately 1 hour	16 (64.0%)
> than 1 hour	3 (12.0%)

Teacher Surveys. Teacher's knowledge, attitudes and practices were measured in a self-report survey, which was administered by trained enumerators to teachers in both control and treatment schools during endline. The survey existed of four sections: teacher's ideas about the learning process, teacher's classroom practices, teacher's motivation, and their opinions and experiences about Catch-Up (only relevant for treatment schools).

Teacher surveys were conducted with the same teachers who participated in the implementation fidelity observations. As mentioned above, when asked if Catch-Up was being implemented at their school, 13 out of 40 observations had a "Not Applicable" or "No" response for this question. Since

a similar question was asked in the Classroom Observations (*"Have you ever had a Catch-Up class before?"),* these observations in the Teacher Survey were compared with observations who responded with "No" to this question in the Classroom Observations during endline[6].

For endline, 7 teachers from treatment schools reported "No" to this question in Classroom Observations. For 2 observations, both teachers during baseline and endline reported "No" to this question. These observations were correspondingly removed from further analysis, as it looks like Catch-Up was not being implemented the way it should in these schools. For 4 schools, no participation in Catch-Up was indicated in both the Implementation Fidelity tool and in the Classroom Observations. Hence, these observations were deleted from the analysis, as it appears that Catch-Up was not being implemented in these schools. In total, 5 observations were removed from data analysis because of Catch-Up implementation appearing to be absent[7]. Nine observations remained with either a "No" or "Not applicable" response in the Teacher Survey to Catch-Up being implemented in their school. These observations were retained for the teacher surveys, as these responses are expected to not be an accurate reflection of reality, since they were not corroborated in other surveys. It is also important to note that these surveys were not about Catch-Up classes specifically, but about teachers' general teaching methods. In addition, removing these observations would reduce the statistical power of the study. The total remaining sample size was 84 schools.

Scores between control and treatment schools were compared. We ran both tabulate commands to look at the absolute differences, as well as tested if these differences were statistically different between control and treatment groups by using one-way ANOVAs. Sample sizes per group (treatment versus control) were consistently higher than 30 and close to equal, hence ANOVA analyses are assumed to be robust for required statistical assumptions. Questions were analysed individually, as they were conceptually diverse.[8]

The first section, called Section A, measures teachers' knowledge about the learning process, looking at classroom practices promoted by Catch Up. This section existed of 11 questions, which were measured on a 4-point Likert scale running from "Strongly disagree" to "Strongly agree".

The first question asked teachers: *"A good teacher demonstrates the correct way to solve a problem."* The difference between control and treatment schools was quite small for this question.

	M	**SD**
Control	3.44	.624
Treatment	3.56	.502

An ANOVA was performed to determine whether the small difference between control and treatment was statistically significant. It revealed that the difference between the means was not statistically significant: (F (1,82) = .92, p = .341). The direct link between this item and Catch-Up promoted practices is not clear. The statement rather refers to a general idea of teaching, since teachers demonstrating incorrect ways of solving problems or not demonstrating how to solve problems at all conflicts with what many would understand a teacher's responsibility to be.

6 For baseline, it is possible that Catch-Up implementation simply had not yet started, so a "No" response during baseline is not considered problematic in and of itself.

7 The following schools were removed: Kaluluzi primary school, Adonai Tildelis, Peniel, Prince Takamado primary, Joseph Conteh. For Adonai Tildelis, absence of Catch-Up implementation was reported in the Implementation Fidelity tool, and in both baseline and endline Classroom Observations, hence why the total number of deleted schools is 5.

8 We decided to not statistically correct for the inflation of a type I error conducting multiple statistical analyses, since the interest of this study is to uncover trends and patterns, more than establishing the statistical significance of certain variables.

The next question asked teachers *"It is better when the teacher – not the learner – decides what activities are done in the classroom"*. As Catch-Up promotes learner-centred classroom activities, treatment schools are expected to score lower on this question than control schools.

	M	**SD**
Control	3.00	.739
Treatment	2.85	.745

As expected, treatment schools score lower on this item than control schools. However, an ANOVA analysis showed that this difference was not statistically significant (F (1,82) = .90, p = .346). The way the question is phrased, could potentially explain why no statistical significance was found. While Catch-Up does promote learners having a certain level of freedom in choosing what they want to do, selection of activities is still mainly teacher directed. Juxtaposing teacher against learner in this question, instead of asking about a collaborative process between the two, could have led Catch-Up teachers to be neutral with regards to this question.

The following item asked teachers *"My role as a teacher is to facilitate the learner's own inquiry"*. This again focused on learner-centred pedagogy, in which the teacher takes the role of the "facilitator" guiding learners through their learning process. Contrary to expectation, control schools reported a higher level of agreement with this statement than treatment schools.

	M	**SD**
Control	3.18	.684
Treatment	2.97	.486

However, an ANOVA analysis showed that this difference was not statistically significant (F (1,82) = 2.40, p = .125).

The following item stated: *"Teachers know a lot more than learners therefore they should not let the learners develop answers that may be incorrect when they can just explain the answers directly."* Teachers at treatment schools were expected to score lower on this than teachers of control schools, since the Catch-Up program promotes learners' freedom to express themselves and the ability to try and find correct answers on their own, with appropriate guidance from teachers. The table below shows that there was no observable difference between treatment and control schools.

	M	**SD**
Control	2.20	.786
Treatment	2.18	.756

An ANOVA analysis confirmed that the small difference between control and treatment schools was not statistically significant (F (1,82) = 0.01, p = .904). The absence of a treatment effect might be at least partially due to the item being quite lengthy, which likely impacted the level of understanding of respondents.

The next item asked teachers *"Learners learn best by finding solutions to problems on their own"*. This item is measuring something conceptually similar to the previous item but phrased in a different way. For this item, teachers at treatment schools are expected to score higher than teachers at control schools. The table below shows they do, but again the difference with control schools is rather small.

	M	SD
Control	2.82	.806
Treatment	2.87	.833

An ANOVA analysis confirmed that the small difference between control and treatment schools was not statistically significant (F (1,82) = 0.08, p = .783). Since this item was phrased in a clearer and more concise way than the previous item, and both measured a similar concept, the absence of large differences between treatment and control schools on this concept cannot only be attributed to unclear wording.

The next item asked teachers what teaching should look like: *"Teaching should be built around problems with clear, correct answers, and around ideas that most learners can grasp quickly"*. Control and treatment schools have similar scores on this item. This makes sense, because the item does not show clear alignment with any of the Catch-Up promoted practices.

	M	SD
Control	3.18	.535
Treatment	3.21	.570

An ANOVA confirms that this difference is not statistically significant (F (1,82) = .05, p = .821).
The next question asked teachers about the importance of learner's background knowledge in the learning process. This relates to the 7 Cs of playful learning, which stresses that teaching should be "connected". Learners need to be able to meaningfully connect what they learn to their existing knowledge and their real lives, to promote understanding of what they learn and why. Based on this background knowledge, a higher score was expected for treatment schools than control schools. The item asked: *"How much learners learn depends on how much background knowledge they have – that is why teaching factors is so necessary"*. The item was quite long and might seem contradictory when one reads it at first, which might have resulted into a small difference between control and treatment group in the opposite direction as expected, as can be seen in the table below.

	M	SD
Control	2.84	.706
Treatment	2.77	.742

An ANOVA determined this difference was not statistically significant (F (1,82) = .23, p = .636).
The following item stated: *"Learners should be allowed to think of solutions to practical problems themselves, before the teacher shows them how they are solved."* This item is related to the previous 2 items asking teachers about learners' freedom to express themselves and the ability to try and find correct answers on their own. Teachers from treatment schools were expected to score higher on this question than teachers from control schools. However, in accordance with results on the previous 2 items, no clear difference was found between treatment and control schools, as can be seen in the table below.

	M	SD

Control	3.22	.704
Treatment	3.23	.627

An ANOVA analysis confirmed that this small difference was not statistically significant (F (1,82) = .00, p = .954).

The following question asked teachers *"When referring to a 'good performance', I mean a performance that lies above the previous achievement of the learner."* This item aimed to measure the idea that learners' individual progress should be the focus when assessing performance, rather than comparing learners amongst each other. Treatment schools are expected to score higher on this item than control schools. However, the table below shows that the small difference between treatment and control schools is in the opposite direction.

	M	**SD**
Control	3.20	.405
Treatment	3.13	.615

An ANOVA analysis confirmed that this small difference was not statistically significant (F (1,82) = .41, p = .524).

The next question asked teachers about the classroom environment. Specifically, it asked: *"A quiet classroom is generally needed for effective learning"*. Learning through play is usually not a quiet activity, because learners actively engage in activities, often together with others. It was thus expected that the treatment group would score lower on this question than the control group.

	M	**SD**
Control	3.18	.806
Treatment	2.79	.864

An ANOVA showed that this difference was statistically significant (F (1,82) = 4.41, p = .039).

The last question in this section asked *"Thinking and reasoning processes are more important than specific curriculum content"*. Catch-Up promotes the development of specific skills sets, rather than using the completion of curriculum content as an indication of learning. In the Zambian context, teachers can be more focused on completing certain curricula and being able to meet national level milestones, instead of cultivating the skills needed to properly understand that curriculum. However, at the same time, teachers cannot only focus on teaching thinking and reasoning skills in the absence of leaners grasping the curriculum content. For this item there is thus a level of ambiguity of what teachers at treatment schools are expected to report. Control schools scored higher on this question than treatment schools, however, this difference was extremely small.

	M	**SD**
Control	3.04	.562
Treatment	2.97	.486

An ANOVA confirmed that indeed this difference was not statistically significant (F (1,82) = .37, p = .546).

In summary, this section did not show a clear pattern of differences between control and treatment schools. Only one question yielded a statistically significant difference, which was the item on classroom environment. All other differences between treatment and control schools were small, sometimes negligible. Of 4 out of 11 items, differences were in the opposite direction of what was expected from a program implementation point of view. Thus, the scores on section A fail to show a difference in teachers in control and treatment schools in terms of teachers' ideas about the learning process.

It is important to note that this section of the teacher survey suffered some problems. Some of the questions were not worded clear enough, which could have limited understanding of respondents, and in turn contributed to the absence of a clear scoring pattern. In addition, some questions were not tied clearly enough to the goals of the Catch-Up program. Also, it could be argued that it would have been better to focus on a specific subset of teaching practices promoted by Catch-Up, and measure those with more than one question. This would have allowed to measure concepts more broadly, capturing more variance in scores, which makes it easier to detect existing population differences. In the current survey, opinions about certain practices promoted by the Catch-Up program were measured with just one question. Indicators measured by individual questions are more prone to random noise in variation that has nothing to do with the concept it is trying to measure. The results of this section should be integrated with other results within this survey, as well as other measurement tools that were used to measure the impact of the program to get a better idea whether differences in teachers' opinions on teaching practices do not exist, or whether problems with the measurement tool led to underestimating these differences.

The next part of the survey, section B, asked teachers to self-report about how they teach in their classrooms. It is important to note that no specific reference to Catch-Up classes was made when asking this set of questions.[9] Questions in this section were measured on a 4-point Likert scale running from "Not at all" to "To a great extent", and the section included 14 questions in total. The average scores and standard deviations of individual questions are reported in the table below. For all these questions, higher scores indicate a higher level of self-reported Catch-Up promoted practices. Similar to the results from the implementation fidelity tool discussed before, we can again see that most scores are above the midpoint of 3, so towards the high end of the scale. This shows that in general teachers report to use Catch-Up practices in their classrooms.

Questions can be mapped to certain components of the 7 Cs conceptual framework, which has been done tentatively in the table below. However, it is important to note that some of these questions can map onto multiple components of the framework, and this mapping can be subjective.

		Control		**Treatment**	
Question	**Mapping on 7C's**	**M**	**SD**	**M**	**SD**
B1. The class includes challenges and puzzles	Cheerful	2.93	.863	3.23	.959
B2. The class includes games, songs, or dances	Cheerful	3.31	.848	3.36	.873
B3. Learners have choices between different activities done in class	Captivating	2.64	1.151	2.38	1.091

[9] The introductory statement of this section read as follows: "In this section I am going to ask you some questions about what you think about how learners learn. Please answer by showing the extent to which the following happens in your classroom."

B4. The class includes hands-on learning experiences	Concrete	3.33	.826	3.56	.788
B5. Learners receive the opportunity to try again if failed	Challenging	3.80	.548	3.71	.674
B6. The teachers give away cues and hints to the students as to help them answering a question	Challenging	3.51	.757	3.18	.790
B7. Learners receive relevant activities in class to the subject taught	Connected	3.58	.811	3.33	.898
B8. I support learners in class using instructional materials adapted to their learning needs	Connected	3.67	.674	3.59	.715
B9. Learners had to work in small groups	Collaborative	3.24	.883	3.54	.854
B10. The class includes positive interactions between the children	Collaborative, Cheerful	3.22	.997	3.54	.720
B11. The class activities include songs and games	Cheerful	3.22	.876	3.41	.785
B12. The class activities are interactive	Concrete	3.67	.707	3.64	.707
B13. Learners are allowed to share their feelings and emotions freely in class	Creative	3.31	.949	2.97	.986
B14. I deliberately make an effort to build relationships with my learners	Collaborative	3.58	.811	3.77	.583

The table above shows some interesting trends. While the overall pattern of differences between treatment and control shows that there is no large difference between them, some patterns appear once we look at the mapping to the components of the 7 Cs framework. In the table above, items for which differences between control and treatment schools were in line with what would be expected based on program implementation are in yellow. Items for which difference between control and treatment schools were in the opposite direction of what was expected are in blue. Items in grey are those items for which differences between treatment and control schools were considered negligible.[10]

For items mapping to the cheerful and collaborative components, treatment school averages tend to be higher than those of control schools, while for items mapping to other components there is no clear pattern.[11] It would be interesting to see if similar trends are found in classroom observations and qualitative interviews.

10 Differences below 0.1 were considered negligible for this comparison.

11 It is important to note that only one item mapped on the components Captivating and Creative, hence no patterns could be determined for these components.

For each item, we also calculated whether differences between control and treatment were statistically significant, which is reported in the table below. All *F*-values are reported with the same degrees of freedom (1,82). The same colour scheme is used as in the previous table, with yellow for expected differences, blue for not expected differences, and grey for negligible differences. None of the differences between control and treatment schools gained statistical significance.

Question	***F*-value**	***p*-value**
B1. The class includes challenges and puzzles	2.24	.139
B2. The class includes games, songs, or dances	0.06	.800
B3. Learners have choices between different activities done in class	1.12	.294
B4. The class includes hands-on learning experiences	1.70	.196
B5. Learners receive the opportunity to try again if failed	0.40	.531
B6. The teachers give away cues and hints to the students as to help them answering a question	3.85	.053
B7. Learners receive relevant activities in class to the subject taught	1.72	.194
B8. I support learners in class using instructional materials adapted to their learning needs	0.26	.614
B9. Learners had to work in small groups	2.39	0.126
B10. The class includes positive interactions between the children	2.70	.0104
B11. The class activities include songs and games	1.06	.307
B12. The class activities are interactive	0.03	.869
B13. Learners are allowed to share their feelings and emotions freely in class	2.54	.115
B14. I deliberately make an effort to build relationships with my learners	1.50	.224

The lack of a general difference between treatment and control schools on this scale could be due to several explanations, which require further investigation. It is possible that there was little impact from program implementation. This conclusion cannot be drawn on these results alone. They will need to be integrated with the results of the Classroom Observations as well as qualitative interviews of teachers. Another important notion is that teachers were asked about their classroom activities in teacher surveys, without a specific reference to Catch-Up classes. There is a possibility that some activities (like group work and signing) might have already been present in teacher classrooms before Catch-Up implementation and did not increase because of Catch-Up implementation in the overall classroom setting. Thus, while the use of these activities might have increased in Catch-Up specific classes, no changes might have been made by teachers in their normal class practices, resulting in an absence of an overall reported change. It is unfortunately not possible to further investigate this suggestion in the absence of a baseline measurement for teacher surveys.

Section C measured teachers' job satisfaction. The section included 7 questions in total, with 6 questions measuring teachers job satisfaction, and one item measuring teachers' impression of learner satisfaction. The first 6 questions in this section were measured on a 4-point Likert scale running from "Strongly disagree" to "Strongly agree", while the last item was measured on a 5-point Likert

scale with the same end point anchors. For all items, higher scores indicated a higher level of satisfaction. Scores of individual items can be found below, with average scores and standard deviations reported.

	Control group		Treatment group	
Questions	**M**	**SD**	**M**	**SD**
C1. All in all, I am satisfied with my job	3.56	.659	3.28	.857
C2. My job helps me to meet my basic needs	3.33	.826	3.10	.718
C3. My job pays me well	2.80	.815	2.38	.815
C4. I consider myself well equipped for teaching	3.49	.549	3.36	.707
C5. I feel that I am making a significant education difference in the lives of my learners	3.62	.535	3.72	.456
C6. I find satisfaction in seeing my learners acquire knowledge in a fun way	3.67	.477	3.69	.468
C7. Learners seem to enjoy the way classes are handled through play	4.47	1.04	4.82	.389

Interestingly, control schools scored higher for items measuring compensation and equipment for the teacher job, while treatment schools scored higher for items measuring contributing towards and enjoying learner progress. While the differences are too small to make any conclusive statements, this is an interesting trend that could be explored in further research. It also resonates with anecdotal evidence of teachers in the qualitative interviews included in deliverable 3. Teachers report that the Catch-Up implementation can be challenging to implement with the limited means they have and that it can be time consuming, which could lead to lower satisfaction in terms of compensation and equipment. However, teachers also describe how the program has helped learners progress.

These results are also in line with another ongoing study on the Catch-Up program. In this study, lack of adequate resources is also reported by teachers as an important barrier to implementing quality Catch-Up classes (Busara, 2023). Teachers in this study do however see the Catch-Up program as an opportunity to learn and grow as a teacher and are motivated by the high participation of learners, similar to results in this study on questions C5 and C6. When it comes to learner satisfaction of the use of learning through play, treatment schools score higher than control schools. Nevertheless, it is important to take this result with a grain of salt. As learning through play was a component of the Catch-Up program, teachers at treatment schools likely have a different understanding of the term 'learning through play' than teachers at control schools. Interestingly, both control and treatment schools report highly positive about the learners' level of enjoyment when participating in learning through play, showing that teachers can easily observe learners' enjoyment when implementing those types of learning methods.

For each item, we also calculated whether differences between control and treatment were statistically significant, which is reported in the table below. All *F*-values are reported with the same degrees of freedom (1,82). Only two items gained significance, as can be seen in the table below.

Questions	***F*-value**	***p*-value**

C1. All in all, I am satisfied with my job	2.73	.103
C2. My job helps me to meet my basic needs	1.84	.179
C3. My job pays me well	5.43	.022
C4. I consider myself well equipped for teaching	0.90	.346
C5. I feel that I am making a significant education difference in the lives of my learners	0.77	.384
C6. I find satisfaction in seeing my learners acquire knowledge in a fun way	0.06	.805
C7. Learners seem to enjoy the way classes are handled through play	4.05	.050

Treatment schools reporting significantly lower levels of satisfaction with their payment is an important result that should be investigated in further research. It would be of specific interest to understand the underlying reasons for that result, to see how that feeling could be mitigated, since dissatisfaction with payment is an important reason for people to quit their jobs (Das & Baruah, 2013; Loeb & Luczak, 2013). The other significantly different score is for the item measuring learner enjoyment of learning through play, which makes sense given that Catch-Up is focused on learning through play, while control schools might or might not have a focus on learning through play methodologies.

The last section, Section D, was relevant only for treatment school respondents, as it asked teachers questions about Catch-Up specifically. In addition, schools for which Catch-Up teachers were not present during the time data was collected, these questions were dropped as well. This resulted in a total sample size of 25 surveys at treatment schools.

The questions in this section focused on teachers' opinions about Catch-Up, as well as the support for Catch-Up they experienced and witnessed within their school. Questions were measured on a 5-point Likert scale running from "Strongly disagree" to "Strongly agree", and a total of 7 questions were included.[12] In the table below we report the values for all questions.

Similar to what was done for the implementation fidelity tool, a *t*-test was performed to see if the score was significantly different from the midpoint assuming data is normally distributed.

Question	**M**	**SD**	***p*-value *t*-test**	**CI *t*-test**
D1. I do not enjoy teaching Catch-Up classes	2.52	1.418	.052	1.935-3.105
D2. Teaching Catch-Up class is not beneficial to my learners	2.00	1.323	.001	1.454-2.546
D3. Other teachers who were also trained on the Catch-Up program are also enjoying teaching the Catch-Up program	3.52	1.122	.015	3.057-3.983
D4. My school authorities do not approve of my participation on the Catch-Up program	2.28	1.458	.011	1.678-2.882
D5. I feel that I do not have the necessary resources to teach the Catch-Up program	2.92	1.441	.392	2.325-3.515

12 Question D8 *"I have not yet started teaching Catch-Up classes"* was removed from analysis, because it should have been measured on a "Yes/No" scale, instead of on a 5-point Likert scale running from "Strongly disagree" to "Strongly agree".

D6. In the face of challenges, I have difficulty to improvise so that I can deliver the Catch-Up class to my learners	2.88	1.394	.335	2.305-3.455
D7. I intend to continue teaching Catch-Up classes for at least 1 year	4.24	.970	.000	3.840-4.640

The average scores here are in the directions that would be expected. Most of the average scores are also significantly different from the midpoint, which is a further confirmation that scores are in the direction as desired by the program. Teachers have positive opinions about the Catch-Up program and have an intention to keep teaching it. They also observe approval from others at the school. For two items scores there was no significant difference between the average score and the midpoint of the answer options. This happened for the items about having the necessary resources and dealing with challenges. This relates to scores on item C4 (*"I considered myself well-equipped for teaching"*), which showed higher scores for control than for treatment school respondents. This also resonates with challenges reported by teachers in the qualitative interviews as well as in ongoing research on Catch-Up implementation barriers and challenges (Busara, 2023).

3 | Conclusion and discussion

In terms of implementation fidelity, high adherence to Catch-Up guidelines was found in observed classrooms. Percentages of adherence to practices ranged from 64% to 88%. For practices that were measured on a quantitative scale, the majority showed values that were significantly higher than the midpoint of the scale. All in all, these results show a high level of implementation fidelity, although adherence to the desired class duration and the use of group leaders to lead classroom activities could still be improved. It is important to reiterate that a relatively large number of observations had to be dropped as a result of Catch-Up teachers not being present and/or Catch-Up classes not being organized on the day of data collection. Future research should allow enough flexibility in timing and budgeting to follow up on schools at which situations like these occur, so they could still be included in observations. The absence of Catch-Up teachers and/or the organization of Catch-Up classes could be an indicator of low-quality Catch-Up implementation, and hence it is possible that the inclusion of these schools could have changed the results.

When it comes to teachers' knowledge and self-reported use of Catch-Up promoted classroom practices, results were inconclusive. Especially for the section measuring knowledge no clear patterns of difference between treatment and control schools was found. Importantly, this section suffered from methodological issues, which have been discussed in more detail above. Future research should establish whether a survey with clearer alignment to Catch-Up, more clearly formulated items, and using scales instead of individual items to measure Catch-Up practices can establish differences between treatment and control schools. The issues with this section also point out the importance of thoroughly pilot testing survey instruments, even when at face value questions seem clear. For the section on knowledge, items mapping to the Cheerful and Collaborative components of the 7Cs framework showed higher scores for treatment schools compared to control schools, while for items mapping on other components there was no clear pattern.

Results being mostly inconclusive in these two sections could be attributed to multiple possible explanations, besides the methodological issues of the section on knowledge. First, there is an expectation that these self-reported questions are specifically prone to socially desirable responding. Teachers at both control schools and treatment schools might be aware of what the desired responses are. The Catch-Up program is likely not the only program in Zambia which focuses on learning through play. The Zambian Ministry of Education has contextualized learning through play in the Zambian context together with the LEGO foundation. This led to the development of the 7Cs framework, providing teachers with a set of guidelines on how to implement learning through play in their classrooms (Ministry of Education Zambia, 2020). This is a national exercise and thus not limited to the scope of treatment schools. In this light, it is particularly important to compare the results of the teacher surveys with the results in the classroom observations, which were conducted by enumerators, and are thus expected to be less prone to social desirability.

Given the development of the 7 Cs framework, it is also possible that there is general progress in Zambian schools in terms of implementing learning through play practices. In addition, there could be spillover effects between the treatment schools and control schools. Even though schools from control and treatment schools were sampled from different provinces, it is possible that teachers still have opportunities to share their learnings with their peers or when being transferred between

schools. Further investigation is needed to determine what causes the lack of a difference between control and treatment schools on teacher surveys. However, it should be noted that while the easiest way to establish program impact would be finding differences between control and treatment schools, the absence of these differences does not necessarily mean that there is a lack of program impact. As some of the explanations above outline, the Zambian government has taken a proactive role into integrating Catch-Up promoted practices into their policies and schools. Thus, the lack of differences could also be an indication of national level progress.

Findings in the section on teacher motivation echoed statements made in another research study on the Catch-Up program conducted in Zambia (Busara, 2023). Positive discoveries are that teachers feel they can make a difference in learners' education, and that learners' participation motivates them. Other research on Catch-Up (Busara, 2023) describes how important these enabling factors are for teachers' quality implementation of the program. Importantly, teachers at treatment schools are less likely to agree that their payment is sufficient than teachers at control schools, and this difference reaches statistical significance. Teachers also report they do not feel they have adequate resources in the section measuring the support they receive for Catch-Up (discussed below), which is also reported in the research by Busara (2023). Both the enabling factors and challenges that are found in this section are important to investigate in future research, as mitigating challenges and promoting enablers can result in sustainable and quality Catch-Up implementation.

Teachers had positive responses when being asked about their opinions on Catch-Up and the support for Catch-Up they experienced in their environment. Most of the scores were also significantly different from the midpoint of the scale, a further confirmation of their positive attitudes. It should however be noted that these questions are also suspected to be prone to socially desirable responding, as they are asking about program implementation specifically. Qualitative interviews in Deliverable 3 will allow for a clearer picture of where these positive attitudes about Catch-Up come from.

Overall, there were only a handful of questions that had statistically significant differences between treatment and control schools. It is important to note that this study had a qualitative and exploratory nature, aiming to discover which patterns and trends currently exist in data. This information can provide useful guidance for designing more robust tests of statistical significance in the future. This exploratory approach was also reflected in the measurement of constructs of interest with individual questions, instead of scales composed of multiple questions. Since scales measure more variance within constructs, they are better designed for tests of statistical significance than individual items are. In addition, reported issues with Catch-Up refresher trainings could have had an influence on the absence of statistical significance. Literature on quality teacher professional development states that a minimum of 20 hours of contact time is considered sufficient to promote intellectual and pedagogical change (Desimone, 2009; Merchie, Tuytens, Devos, & Vanderlinde, 2016). However, when these development activities are organized over a short period of time, it is important that follow ups are organized (Desimone, 2009; Merchie, Tuytens, Devos, & Vanderlinde, 2016). The Catch-Up program organizes their trainings over a course of 4 days, after which refresher trainings take place periodically. Nevertheless, in a study on the Catch-Up program in Zambia some teachers reported they are not invited for these trainings, while others stated it takes a long time before they are organized (Busara, 2023). Across the board, teachers echoed the need for more and frequent refresher trainings (Busara, 2023). These issues concerning refresher trainings could thus hamper teachers' intellectual and pedagogical change, which might explain the absence of desired outcomes.

The current study did detect interesting patterns of differences between treatment and control schools on the use of some of the components of the 7Cs framework in teachers' classroom practices, and on teacher motivation aspects. Furthermore, it found positive attitudes of teachers and their school environments for the Catch-Up program. Results in this deliverable will be integrated with the results of Deliverable 3, to determine whether consistent patterns can be found.

Deliverable III: Classroom Observations and Qualitative Interviews

1 | Methods Classroom Observations

1.1 Participants and Context Classroom Observations

Data was collected as part of a larger scale evaluation of the implementation of the Catch-Up program in Lusaka province in Zambia. The evaluation aimed to determine both learners' social and emotional learning (Deliverable 1), as well as the impact of the program on teacher's knowledge, attitudes and practices (Deliverable 2) and spill-over of Catch-Up promoted classroom practices in regular classes (current deliverable). The current study conducted classroom observations at both treatment and control schools, at both baseline and endline. In additional to classroom observations, qualitative interviews were conducted with 30 teachers, as well as 5 stakeholders (government officials and program implementation staff), to get an in-depth understanding of how, when, and why Catch-Up promoted practices are implemented in regular classes. Below, the design, data collection, and results of the classroom observations are described first, after which the design, data collection, and results of the qualitative interviews are described. Results of both classroom observations and qualitative interviews are then summarized in a general discussion section.

1.2 Design and Randomization Classroom Observations

Data was collected in 3 districts in Lusaka province and 2 districts in Central province.[13] The Catch-Up program was implemented in Lusaka province, hence only treatment schools were sampled from Lusaka. Catch-Up was not implemented in Central province, so all schools selected from there served as control schools. Classroom observations were conducted at 40 control and 46 treatment schools during both baseline and endline, resulting in a total sample size of 86 schools. One teacher and one regular class were observed per data collection moment, per school. This resulted in a total of 2 observations per school: one during baseline and one during endline. Schools and classes sampled for baseline and endline were always the same, however teachers who were observed during baseline were not always the same teachers as observed during endline. This occurred because Catch-Up implementation happens at the class level and not the teacher level. Thus, if a teacher changed classes or transferred to another school, Catch-Up implementation would be taken over by the new teacher assigned to that classroom. Schools were selected from districts in Lusaka province that were new to Catch Up. Only classrooms within which Catch-Up was implemented were selected, and hence only teachers who were expected to be trained in Catch-Up were invited to participate.[14] Only one

13 The districts were Chilanga, Luangwa, Lusaka, Chibombo, and Shibuyunji.

14 However, it is possible that teachers who were only recently assigned to Catch-Up classes at the time of endline data collection might not have yet been trained on Catch-Up.

classroom and thus only one teacher was invited to participate per school. Teachers who participated in the classroom observations were the same teachers who participated in the implementation fidelity observations and teacher surveys from Deliverable 2.

1.3 Assessment of Constructs Classroom Observations

Classroom Observations were conducted by trained enumerators in regular classes (not Catch-Up classes), to measure whether there is a spillover of Catch-Up practices into general teaching practices. They reported what they observed using a standardized tool. This was a newly developed tool, to assure that it was sufficiently contextualized to the implementation of Catch-Up in Zambia. Questions were developed based on the theory of the 7 Cs framework, written by the Zambian Ministry of Education. Question development was a collaborative and iterative process between CAPOLSA and VVOB, with multiple feedback rounds. As part of this process, the CAPOLSA team attended a VVOB administered Catch-Up training, to allow for a deeper understanding of the constructs to be measured. After tool development, a piloting exercise was conducted at 2 schools in Lusaka province which would not be included in the baseline and endline. The piloting exercise allowed the research team to understand which items were not properly understood, and these were adjusted accordingly.

The final tool existed of one section that was filled in while the enumerator was observing the class, which measured the presence of general Catch-Up promoted practices. Afterwards, enumerators filled out a second section which included questions aiming to measure the presence of the components of the 7Cs framework in the observed teaching practices.

The first section existed of 15 questions. The first 2 questions were measured on a "Yes/No" scale and asked, *"The teacher has a lesson plan"* and *"The teacher sings songs with the children"*. The remaining 13 questions were all measured on a 5-point Likert scale running from "Never" to "Very often". The questions showed high internal consistency (Cronbach's a = .8816). Example items are *"The teacher encourages the children to take the lead in small groups or class"* and *"The teacher encourages the child to find the answer to the question"*.

The second section existed of 7 sets of questions, with every set aiming to measure one of the 7 components of the 7 Cs framework. All questions were measured on a 5-point Likert scale running from "Not at all" to "Very much". Each set of questions is discussed individually.

Cheerful. This component refers to teaching activities that learners enjoy engaging in. It was measured with a set of 3 questions, which had sufficient internal consistency to be combined in a scale score (Cronbach's a = .8141). An example item is *"The class included games, songs or dances."*

Captivating. This component refers to teaching activities that attract and hold a learners' interest and allow the learner to make choices about the learning process. It was measured with a set of 5 questions, which had sufficient internal consistency to be combined in a scale score (Cronbach's a = .8739). An example item is *"Learners had choices between different activities done in class"*.

Challenging. This component refers to teaching activities that allow learners to discover things for themselves, instead of being provided with readymade solutions. It was measured with a set of 4 questions, which had sufficient internal consistency to be combined in a scale score (Cronbach's a = .8881). An example item is *"Learners received the opportunity to try again if they failed"*.

Connected. This component refers to teaching activities that relate to something already known by the learner, as well as being connected to their needs and interests. It was measured with a set of 6 questions, which had sufficient internal consistency to be combined in a scale score (Cronbach's a = .8833). An example item is *"The teacher refers to the local context of the learners in his/her lessons"*.

Collaborative. This component refers to teaching activities that allow learners to interact and collaborate with their peers. It was measured with a set of 4 questions, which had sufficient internal consistency to be combined in a scale score (Cronbach's a = .8195). An example item is *"Learners had to work in small groups"*.

Concrete. This component refers to teaching activities that provide learners with the opportunity to actively engage with, manipulate and transform materials. It was measured with a set of 4 questions, which had sufficient internal consistency to be combined in a scale score (Cronbach's a = .8039). An example item is *"The teacher uses concrete materials such as stones, sticks, baskets, etc."*.

Creative. This component refers to teaching activities that give learners the possibility to express their thoughts and emotions freely, without restrictions imposed by the teacher. It was measured with a set of 3 questions, which had sufficient internal consistency to be combined in a scale score (Cronbach's a = .8115). An example item is *"The classroom environment was open and comfortable"*.

1.4 Data Collection Classroom Observations

Approval to conduct the study was obtained from the Ministry of Education, and ethical clearance was obtained from the Research Ethics Committee of the University of Zambia. Baseline data collection took place in February 2022, three weeks after the rollout of the Catch-Up program in Lusaka province. Endline data collection took place in September-November 2022, roughly 7-9 months after the start of the Catch-Up program. Both baseline and endline data was collected by the Centre for the Promotion of Literacy in Sub-Saharan Africa (CAPOLSA). Classroom observations were filled out by trained enumerators themselves, as they observed a regular, non-Catch-Up class. All respondents participating in the study were asked for informed consent. Participation was voluntary and a respondent could always refuse collaboration during the survey. All data was collected on the Kobo collect app. Data was downloaded in Excel format and converted to Stata datasets for data cleaning and analysis.

2 | Empirical Strategy and Results Classroom Observations

The classroom observation data was analysed using Stata 18.

As mentioned above, classroom observations were conducted at both baseline and endline, at both treatment and control schools. Scores between control and treatment schools were compared. We both looked at the absolute differences, as well as tested if these differences were statistically different between control and treatment groups by using one-way ANCOVAs, including baseline scores as covariates. Correlating for baseline scores allows us to consider teachers' starting point, correcting for unobservable but existing differences between teachers that might exist before program implementation took place and that impact outcome variables of interest. This gives us the opportunity to also detect smaller but relevant changes in our outcome variables of interest. Sample sizes per group (treatment versus control) were consistently higher than 30 and close to equal, hence AN(C)OVA analyses are assumed to be robust for required statistical assumptions.[15]

Questions which were intended to measure a similar concept and showed high internal consistency were summarised in scale scores. These were computed by summing the scores of items in one scale and dividing this sum by the total number of items included.

For some schools included in the dataset, data was collected at only one timepoint (either baseline or endline). This was the case for 11 schools. Since the ANCOVA analysis requires scores for both baseline and endline, these observations were dropped from further analysis. Teachers observed were asked if they had ever had a Catch-Up class before. For baseline, 10 teachers from treatment schools reported they had not. This can be explained by the baseline taking place when some schools had not yet started with Catch-Up implementation. For endline, 7 teachers from treatment schools reported "No" to this question. For 2 observations, both teachers during baseline and endline reported "No" to this question. These observations were correspondingly removed from further analysis, as it looks like Catch-Up was not being implemented in these schools. For 5 observations, teachers at endline were different from who was interviewed during baseline. Teachers who were interviewed during baseline for these 5 schools did report they participated in Catch-Up activities. It is thus possible that the teachers who participated in the endline measurement were not (yet) participating in Catch-Up at the time we interviewed them, however, the school at large is expected to participate in the Catch-Up implementation. Since a similar question was asked in the Implementation Fidelity tool *("Is Catch-Up being implemented at your school currently?"*), observations who reported no participation in Catch-Up were compared with observations who reported no participation in Catch-Up in the Implementation Fidelity tool. For 4 schools, no participation in Catch-Up was indicated in both the Implementation Fidelity tool and in the Classroom Observations. Hence, these observations were deleted from the analysis, as it appears that Catch-Up was not being implemented in these schools. In total, 5

15 We decided to not statistically correct for the inflation of a type I error conducting multiple statistical analyses, since the interest of this study is to uncover trends and patterns, more than establishing the statistical significance of certain variables.

observations were removed from data analysis because of Catch-Up implementation appearing to be absent[16]. The total remaining sample size was 77 schools.

The first section of the survey measured general implementation of Catch-Up promoted practices in regular classes. At endline this section included 15 questions. The first 2 items on the endline scale were measured on a "Yes"/"No" scale. These items were not measured (in the same way) during the baseline[17]. Hence, for these questions descriptive statistics were performed comparing control and treatment schools, without controlling for baseline scores.

	Control		**Treatment**	
Question	**Yes**	**No**	**Yes**	**No**
The teacher has a lesson plan	24 (66.7%)	12 (33.3%)	30 (73.2%)	11 (26.8%)
The teacher sing songs with the children	3 (8.3%)	33 (91.7%)	14 (34.2%)	27 (65.9%)

The table shows that for treatment schools enumerators observe a higher percentage of teachers who have lesson plans, and who sing songs with the children compared to control schools, which is in line with what is promoted in Catch-Up. Chi-square tests were used to determine whether these differences in percentages were statistically significant. For the first question, differences between control and treatment schools were not significant $\chi^2(1) = 0.39$, $p = .534$. For the second question, differences between control and treatment schools were significant: $\chi^2(1) = 7.42$, $p = .006$.

The remaining 13 questions in the first section were measured on a 5-point Likert scale running from "Never" to "Very often". Items had high internal consistency at endline (Cronbach's $a = .8797$), and they were created based on the theory of the 7C's framework. Thus, all items are expected to measure similar concepts, hence a scale score was created[18]. Treatment schools score higher than control schools on the frequency of Catch-Up promoted activities they implement, as shown in the table below.

	M	**SD**
Control	2.02	.574
Treatment	2.34	.831

An ANOVA showed that this difference was close to significance ($F(1,73) = 3.75$, $p = .057$). When correcting for the average baseline score in an ANCOVA[19], the difference between treatment and control schools becomes statistically significant. This shows that at endline, teachers at treatment schools are observed to implement Catch-Up practices more frequently in their teaching than teachers at control schools.

Article I.	**Source**	**Article II. f**	**D**	***Article III.***	***F***	**Article IV. value**	***p-***
Article V.		Article VI.		Article VII.		Article VIII.	
Article IX.	***Model***	Article X.	2	Article XI. 5	2.4	Article XII. 4	0.09

16 The following schools were removed: Kaluluzi primary school, Adonai Tildelis, Peniel, Prince Takamado primary, Joseph Conteh. For Adonai Tildelis, absence of Catch-Up implementation was reported in the Implementation Fidelity tool, and in both baseline and endline Classroom Observations, hence why the total number of deleted schools is 5.

17 G1_EL *"The teacher has a lesson plan"* was not measured during baseline. G2_EL *"The teacher sing songs with the children"* was measured on a 5-point Likert scale running from "Never" to "Very often" during baseline.

18 Internal consistency for the same 13 items during baseline was also high, with Cronbach's $a = .8772$. Items were thus combined in a scale score. The baseline Classroom Observation tool can be found in Annex 4.

19 Complete ANCOVA tables are reported in Annex 3.

Article XIII.		Article XIV.		Article XV.		Article XVI.	
Article XVII.	***Treatment***	Article XVIII.	1	Article XIX.	4.90	Article XX.	0.03
Article XXI.	***Baseline score***	Article XXII.	1	Article XXIII.	0.50	Article XXIV.	0.48
Article XXV.		Article XXVI.		Article XXVII.		Article XXVIII.	

We also looked at the differences between treatment and control schools for individual questions, to see if there were any observable trends for questions which measure similar constructs. The table below shows the mean scores for both control and treatment schools for every question, as well as reports the *p*-value and confidence interval for the ANOVA testing if differences in means were statistically significant.

	Control group		**Treatment group**		**Significance test difference**	
Question	**M**	**SD**	**M**	**SD**	***p*-value**	**CI of the difference between the means**
G3. The teacher dances with the children	1.11	.398	1.49	.952	.030	-.716 / -.037
G4. The teacher plays word games with the children	1.28	.741	1.90	1.28	.012	-1.109 / -.141
G5. The teacher plays number games with the children	1.19	.710	2.00	1.37	.003	-1.319 / -.292
G6. The teacher organises small group activities among the children	1.28	.88	1.68	1.035	.071	-.845 / -.035
G7. The teacher encourages the children to tell a story to the other children in small groups or class	1.19	.624	1.73	1.28	.025	-1.006 / -.068
G8. The teacher encourages the children to take the lead in small groups or class	1.72	1.059	2.07	1.273	.196	-.887 / .185
G9. The teacher encourages the child to find the answer to the questions	3.19	.980	3.24	1.135	.840	-.534 / .435
G10. The teacher encourages the child to ask another child the answer on a question when he/she does not know the answer	1.83	1.183	2.12	1.288	.312	-.853 / .276
G11. The teacher creates a welcoming atmosphere for all children	3.36	1.268	3.41	1.245	.853	-.625 / .518
G12. The teacher talks with the children in the mother tongue	3.64	1.334	3.71	1.289	.820	-.665 / .528
G13. The teacher talks with the children on things that are not related to a subject or exam	1.53	.910	1.62	1.21	.725	-.582 / .407
G14. The teacher encourages the children to express his/her feelings	2.39	1.153	2.37	1.240	.933	-.523 / .569
G15. The children laugh in class	2.56	1.229	2.76	1.200	.472	.753 / .352

For the individual questions, treatment schools score consistently higher than control schools, with only two exceptions. Since this section measures use of Catch-Up promoted practices, with higher scores reflecting more frequent use of practices, this is in line with expectation. Statistically significant

differences are found for 4 questions, with treatment schools scoring higher than control schools. These questions are specifically focused on the use of playful activities in class, including dances, games, and stories. It would be interesting to see if this finding can be further corroborated in other sections of the Classroom Observations, or in the qualitative interviews.

The second section of the survey was filled out by enumerators after observing a regular class and focused specifically on the 7 Cs framework. All questions were answered on a 5-point Likert scale, running from "Not at all" to "Very much". Each of the components is discussed separately below, starting with the component of Cheerful. Questions measuring these components were designed based on the theory of the 7 Cs framework and were thus considered to be conceptually similar.

Cheerful.

This component was measured with a set of 3 questions at endline, which had sufficient internal consistency to be combined in a scale score (Cronbach's a = .8141).[20] Individual question scores per question, as well as the computed scale score can be found below.

Question	**Control group**		**Treatment group**	
	M	**SD**	**M**	**SD**
Scale score Cheerful	1.70	.73	2.10	1.08
The class included challenges and puzzles	1.69	.920	2.10	1.281
The class included games, songs or dances	1.33	.894	1.78	1.255
The class contained activities that asked from the children to use all senses	2.08	.874	2.41	1.224

In line with expectation, treatment schools scored higher than control schools for each of these questions, as well as for the general scale score. An ANOVA showed that the difference on the cheerful scale score was close to significance (F (1,75) = 3.40, p = .069). Correcting for the average baseline score in an ANCOVA, did not affect the difference between treatment and control schools much, as can be seen in the table below.

Article XXIX.	**Source**	**Article XXX.**	**Df**	***Article XXXI.***	***F***	**Article XXXII.**	***p*-value**
Article XXXIII.		Article XXXIV.		Article XXXV.		Article XXXVI.	
Article XXXVII.	***Model***	Article XXXVIII.		Article XXXIX.	1.82	Article XL.	0.169
Article XLI.		Article XLII.		Article XLIII.		Article XLIV.	
Article XLV.	***Treatment***	Article XLVI.	1	Article XLVII.	3.64	Article XLVIII.	0.060
Article XLIX.	***Baseline score***	Article L.	1	Article LI.	0.28	Article LII.	0.601
Article LIII.		Article LIV.		Article LV.		Article LVI.	

Captivating.

This component was measured with a set of 5 questions, which had sufficient internal consistency to

20 At baseline, this component was measured with the same 3 questions, with an internal consistency of Cronbach's a = 6205. These items were combined in a scale score, which was used as the covariate in the ANCOVA for the Cheerful component.

be combined in a scale score (Cronbach's a = .8739).[21] Scores on both the computed scale as well as individual questions are shown in the table below.

Question	**Control group**		**Treatment group**	
	M	**SD**	**M**	**SD**
Scale score Captivating	2.60	.861	2.94	.957
Learners engage and participate in classroom activities	3.19	1.191	3.51	1.227
Learners show interest in classroom activities	3.08	1.228	3.46	1.247
Learners had choices between different activities done in class	1.31	.710	1.51	.810
The class included hands-on learning experiences	1.89	1.063	2.51	1.287
The teacher asked questions to the students	3.53	1.133	3.71	1.167

Also for the captivating component, treatment schools score consistently higher than control schools, as expected by program implementation. An ANOVA was performed to determine if the difference between treatment and control schools was significant for the overall scale score, but significance could not be established: (F (1,75) = 2.68, p = .106). Correcting for baseline scores in an ANCOVA did not change this, as can be seen below:

Article LVII. Source	**Article LVIII. Df**	***Article LIX. F***	**Article LX. *p*-value**
Article LXI.	Article LXII.	Article LXIII.	Article LXIV.
Article LXV. Model	Article LXVI. 2	Article LXVII. 1 .33	Article LXVIII. 0. 271
Article LXIX.	Article LXX.	Article LXXI.	Article LXXII.
Article LXXIII. Treatment	Article LXXIV. 1	Article LXXV. 2 .53	Article LXXVI. 0. 116
Article LXXVII. Baseline score	Article LXXVIII.	Article LXXIX. 0 .01	Article LXXX. 0. 909
Article LXXXI.	Article LXXXII.	Article LXXXIII.	Article LXXXIV.

Challenging.

This component was measured with a set of 4 questions, which had sufficient internal consistency to be combined in a scale score (Cronbach's a = .8881).[22] Scores on the combined scale score, as well as individual scores are reported below:

Question	**Control group**		**Treatment group**	
	M	**SD**	**M**	**SD**
Scale score Challenging	2.78	1.04	2.87	1.22
Learners help each other to solve challenges and complete tasks given by teachers	2.14	1.291	2.51	1.340

21 At baseline, this component was measured with just 3 out of the 5 questions included during endline. The internal consistency of these items was not sufficiently high to combine them in a scale score (Cronbach's a = .3719). This appeared to be solely driven by the *item "The teacher asked questions to the students"*, hence this item was removed when computing the scale score. The remaining 2 items had a Cronbach's a = .6286, and thus were combined in a scale score. This scale score was used as the covariate in the ANCOVA for the Captivating component.

22 At baseline this component was measured by 3 out of 4 questions used at endline. These questions had an internal consistency of Cronbach's a = .8210. These items were combined in a scale score, which was used as the covariate in the ANCOVA for the Challenging component.

Learners received the opportunity to try again if failed	2.61	1.358	2.78	1.525
The teachers gave away cues and hints to the students as to help them answering a question	3.08	1.251	3.02	1.313
The teacher helped students by giving them appropriate ways of solving a question	3.31	1.091	3.15	1.216

For the overall scale score, treatment schools score higher than control schools as expected. However, when looking at the individual questions, only the first two questions have the same pattern. For the third question, the difference between control and treatment schools seems negligible. For the last question, the difference between control and treatment schools is in the opposite direction as would be expected from a program implementation standpoint. The ANOVA performed to check for the significance of the difference between treatment and control schools on the scale score was not significant (F (1,75) = .10, p = .756). Correcting for baseline scores in a supplemental ANCOVA led to similar results:

Article LXXXV.	**S**	**Article LXXXVI.**		***Article LXXXVII.***		**Article LXXXVIII.**	
ource		**f**				**-value**	
Article LXXXIX.		Article XC.		Article XCI.		Article XCII.	
Article XCIII.	***Model***	Article XCIV.	2	Article XCV.	0.5	Article XCVI.	0.60
				0		7	
Article XCVII.		Article XCVIII.		Article XCIX.		Article C.	
Article CI.	***Treatm***	Article CII.	1	Article CIII.	0.3	Article CIV.	0.57
ent				2		0	
Article CV.	***Baselin***	Article CVI.	1	Article CVII.	0.9	Article CVIII.	0.34
e score				1		4	
Article CIX.		Article CX.		Article CXI.		Article CXII.	

Connected.

This component was measured with a set of 6 questions, which had sufficient internal consistency to be combined in a scale score (Cronbach's a = .8833).[23] Both the scale score and the scores for individual questions are reported in the table below:

Question	**Control group**		**Treatment group**	
	M	**SD**	**M**	**SD**
Scale score Connected	2.94	1.01	2.98	1.12
The teacher refers to the local context of the learners in his/her lessons	2.92	1.442	2.88	1.400
The teacher uses examples from real life or from previous lessons	2.92	1.500	2.90	1.463
Learners received relevant activities in class to the subject taught	2.92	1.273	2.95	1.396
The teacher used templates, structures, or examples in class from previous lessons	2.50	1.231	2.63	1.462
The teacher supported learners in class using instruction materials adapted to the learning needs of the learners	2.69	1.283	2.59	1.245

23 At baseline this component was measured by 3 out of 6 questions used at endline. These questions had an internal consistency of Cronbach's a = 8388. These items were combined in a scale score, which was used as the covariate in the ANCOVA for the Connected component.

The teacher uses familiar language in his/her interaction with the learners	3.67	1.309	3.93	1.058

This section shows very small, nearly negligible differences for the scale score and half of the individual questions. Only the last 3 questions have larger differences between treatment and control schools. These differences were in the direction as expected for 2 items, with treatment schools having higher scores than control schools: "*The teacher used templates, structures, or examples in class from previous lessons"* and *"The teacher uses familiar language in his/her interaction with the learners"*. For 1 item, control schools scored higher than treatment schools, which was for the item *"The teacher supported learners in class using instruction materials adapted to the learning needs of the learners"*. All in all, it looks like for this section there are no observable differences between control and treatment schools. This was confirmed by an ANOVA, showing that there were no significant differences between treatment and control schools on the scale score for the Connected component: (F (1,75) = .03, p = .856). When correcting for the average baseline score on the Connected scale by performing an ANCOVA, the difference between treatment and control schools remained not significant.

Article CXIII. Source	**Article CXIV. Df**	***Article CXV. I***	**Article CXVI. *p*-value**
Article CXVII.	Article CXVIII.	Article CXIX.	Article CXX.
Article CXXI. Model	Article CXXII. 2	Article CXXIII. 0 .11	Article CXXIV. 0.899
Article CXXV.	Article CXXVI.	Article CXXVII.	Article CXXVIII.
Article CXXIX. Treatment	Article CXXX. 1	Article CXXXI. 0 .11	Article CXXXII.0.741
Article CXXXIII. Baseline score	Article CXXXIV.	Article CXXXV.0 .05	Article CXXXVI. (.826
Article CXXXVII.	Article CXXXVIII.	Article CXXXIX.	Article CXL.

Collaborative.

This component was measured with a set of 4 questions, which had sufficient internal consistency to be combined in a scale score (Cronbach's a = .8195).[24] The scale score and individual question scores can be found in the table below.

Question	**Control group**		**Treatment group**	
	M	**SD**	**M**	**SD**
Scale score Collaborative	2.21	.950	2.60	1.16
Students had to work in small groups	1.39	1.050	1.80	1.308
The class included positive interactions between children	2.58	1.317	3.00	1.304
The class included positive interactions between the teacher and the children	3.14	1.400	3.32	1.312
Students had to present their work to the other children in class	1.72	1.279	2.29	1.569

For the collaborative component, treatment schools consistently score higher than control schools on both the overall scale score and all individual questions, albeit that the difference on the second

24 At baseline this component was measured by the same 4 questions as at endline. These questions had an internal consistency of Cronbach's a = .7694. These items were combined in a scale score, which was used as the covariate in the ANCOVA for the Collaborative component.

item (*"The class included positive interactions between children"*) is negligible. However, an ANOVA showed that the difference between control and treatment schools on the scale score is not statistically significant (F (1,75) = 2.65, p = .108). An ANCOVA correcting for baseline scores on the collaborative component, confirmed this[25].

Article CXLI. Source	**Article CXLII. Df**	***Article CXLIII. F***	**Article CXLIV. *p*-value**
Article CXLV.	Article CXLVI.	Article CXLVII.	Article CXLVIII.
Article CXLIX. Model	Article CL. 2	Article CLI. 3.69	Article CLII. 0.030
Article CLIII.	Article CLIV.	Article CLV.	Article CLVI.
Article CLVII. Treatment	Article CLVIII. 1	Article CLIX. 1.72	Article CLX. 0.193
Article CLXI. Baseline score	Article CLXII. 1	Article CLXIII. 4.61	Article CLXIV. 0.035
Article CLXV.	Article CLXVI.	Article CLXVII.	Article CLXVIII.

Concrete.

This component was measured with a set of 4 questions, which had sufficient internal consistency to be combined in a scale score (Cronbach's a = .8039).[26] The table below shows both the combined scale score as well as the scores for individual questions.

Question	**Control group**		**Treatment group**	
	M	**SD**	**M**	**SD**
Scale score Concrete	1.61	.733	2.08	1.057
The teacher uses concrete materials such as stones, bricks, baskets, etc	1.28	.779	1.71	1.230
The class activities included songs and games	1.36	.899	1.78	1.215
The class included story telling activities	1.31	.889	1.80	1.400
The classroom activities were interactive	2.50	1.404	3.02	1.313

This component shows consistently higher scores for treatment school responses than for control schools, for both the overall scale scare as all individual questions. An ANOVA computing the significance of the difference between treatment and control schools on the overall scale score, showed that this difference is significant: (F (1,75) = 4.96, p = .029). This significance remains when correcting for baseline score on the concrete component scale in an ANCOVA, as can be seen below.

Article CLXIX. Source	**Article CLXX. Df**	***Article CLXXI. F***	**Article CLXXII. *p*-value**
Article CLXXIII.	Article CLXXIV.	Article CLXXV.	Article CLXXVI.
Article CLXXVII. Model	Article CLXXVIII.	Article CLXXIX. .29	Article CLXXX. 0.043
Article CLXXXI.	Article CLXXXII.	Article CLXXXIII.	Article CLXXXIV.

25 This ANCOVA did show a significant impact of the covariate baseline score. For baseline, treatment schools (M = 2.31, SD = .811) scored higher than control schools (M = 2.59, SD = 1.009). However, also for baseline this difference was not statistically significant, as confirmed by an ANOVA analysis: (F (1,75) = 1.84, p = 179.).

26 At baseline this component was measured by 3 out of 4 questions used at endline. These questions had an internal consistency of Cronbach's a = .6641. These items were combined in a scale score, which was used as the covariate in the ANCOVA for the Concrete component.

Article CLXXXV. Treatment	***T***	Article CLXXXVI.	Article CLXXXVII. .06	Article CLXXXVIII. .048
Article CLXXXIX. Baseline score	***B***	Article CXC. 1	Article CXCI. 1.58	Article CXCII. 0.21 3
Article CXCIII.		Article CXCIV.	Article CXCV.	Article CXCVI.

Creative.

This component was measured with a set of 3 questions, which had sufficient internal consistency to be combined in a scale score (Cronbach's *a* =.8115).[27] The table below displays scores for the combined scale score and the individual questions.

Question	**Control group**		**Treatment group**	
	M	**SD**	**M**	**SD**
Scale score Creative	2.56	1.089	2.79	1.089
Learners were allowed to share their feelings and emotions freely	2.19	1.238	2.39	1.159
The classroom environment was open and comfortable	2.53	1.276	3.02	1.214
The teacher made an effort to build relationships with learners	2.97	1.362	2.95	1.396

For this component, again treatment schools scored consistently higher than control schools as expected, with the exception of the last item, for which the difference looks negligible. An ANOVA showed that the difference between treatment and control schools was however, not significant (*F* (1,75) = 0.81, *p* = .371). This was also confirmed by an ANCOVA, correcting for baseline scores on the Creative scale, as can be seen below.

Article CXCVII. Source	**Article CXCVIII. f**	***Article CXCIX. F***	**Article CC. *p*-value**
Article CCI.	Article CCII.	Article CCIII.	Article CCIV.
Article CCV. Model	Article CCVI. 2	Article CCVII. 6.18	Article CCVIII. 0.003
Article CCIX.	Article CCX.	Article CCXI.	Article CCXII.
Article CCXIII. Treatment	Article CCXIV. 1	Article CCXV. 0.01	Article CCXVI. 0.930
Article CCXVII. Baseline score	Article CCXVIII.	Article CCXIX. 11.45	Article CCXX. 0.001
Article CCXXI.	Article CCXXII.	Article CCXXIII.	Article CCXXIV.

As can be seen in this table, the baseline score did significantly predict endline scores. In other words, the higher someone scored on the baseline Creative scale, the higher someone scored on the endline Creative scale. Similar to endline, treatment schools (*M* = 3.18, *SD* = 1.138) scored higher than control schools (*M* = 2.56, *SD* = .809) on the Creative scale. However, for baseline this difference was statistically significant, as confirmed by an ANOVA analysis: (*F* (1,75) = 7.26, *p* = .009). These tests show that while the difference between treatment and endline schools is in the expected direction of the program, this difference has reduced instead of increased from baseline to endline.

27 At baseline this component was measured by the same questions as during endline. These questions had an internal consistency of Cronbach's *a* = 9088. These items were combined in a scale score, which was used as the covariate in the ANCOVA for the Creative component.

3 | Methods Qualitative Interviews

3.1 Participants and Context Qualitative Interviews

Besides the classroom observations discussed above, the current study also conducted qualitative interviews. Thirty teachers, as well as 5 stakeholders (government officials and program implementation staff), participated. The aim of the interviews was to get an in-depth understanding of how, when and why Catch-Up promoted practices are implemented in regular classes, with a specific focus on elements of learning through play. Interviews were conducted through the phone and were recorded. Recordings were transcribed, and full transcriptions were used for data analysis. A thematic analysis approach was used.

3.2 Design and Randomization Qualitative Interviews

Only teachers from treatment schools were selected to participate. As mentioned above, schools were selected from three districts in Lusaka province that were new to Catch Up. Per school only one teacher was sampled. The same teachers who were selected to participate in classroom observations, teacher surveys, and implementation fidelity observations were invited to participate. Some of the teachers could not be reached through phone, resulting in a total sample size of 30 teachers. Of the teachers who participated, 5 were head teachers and 25 were regular teachers.

As for stakeholders, government officials were selected from Lusaka province, since all treatment schools sampled for this study were selected from there. Three District Resource Center Coordinators (DRCCs) were included, as well as one Education Officer Teacher Education (EOTEd). The program implementation staff was an Education Advisor from VVOB, who besides implementing the Catch-Up program in Lusaka also had experience implementing the program in other parts of Zambia.

3.3 Assessment of Constructs Qualitative Interviews

Question development of the qualitative interviews drew upon the paper of Kaymakamoğlu (2018). The author of this paper based their question development on a comparison between traditional (teacher-centred) versus constructivist (learner-centred) models of learning. This comparison between traditional and constructivist models of learning showed a clear alignment with the theory outlined in the 7 Cs framework of learning through play. While Kaymakamoğlu (2018) looked at 10 different dimensions along which traditional and constructivist models were compared, the current study focused on dimensions of the teacher and learner roles in the classroom, learning experiences, and views on learning and knowledge, as these were considered the most relevant to the Catch-Up program. The questions developed were adjusted to the Catch-Up context. Probes were included in the interview guide that asked respondents specifically about the use of teaching methodologies that reflected characteristics of the 7 Cs framework of learning through play. The interview guide can be found in Annex 4.

3.4 Data Collection Qualitative Interviews

Approval to conduct the study was obtained from the Ministry of Education, and ethical clearance was obtained from the Research Ethics Committee of the University of Zambia. Interviews were only conducted at endline, in September-November 2022. This was roughly 7-9 months after the start of the Catch-Up program. This data collection was also conducted by the Centre for the Promotion of Literacy in Sub-Saharan Africa (CAPOLSA). Interviews were conducted through the phone and were recorded. All respondents participating in the study were asked for informed consent. Participation was voluntary and a respondent could always refuse collaboration during the survey. Interviews were transcribed and transcripts shared with the main researcher for analysis.

3.5 Empirical Strategy and Results Qualitative Interviews

The qualitative data of the interviews with teachers and stakeholders was analysed using thematic analysis. No software was used during data coding or analysis. Interviews had one section with questions about teachers, and one section with questions about learners. Every section contained one main question and a predetermined set of follow up questions. The findings are summarised per question.

Main question teacher: From your experience, what do you think is the role of the teacher in a classroom setting?

For teachers, two roles were most salient. Half of the interviewees mentioned that a teacher has the responsibility to impart knowledge on learners, and to assure that learning goals are met. This idea is illustrated by the answers below:

"So the role of a teacher in a classroom setting is to make sure the children are taught and get the concept the teacher is teaching."

"It is to make sure that the learners are able to read and write."

In addition to imparting knowledge on learners, teachers often view themselves as instructors or facilitators. This role is seen as someone who provides the learners with instructions about what they are supposed to be doing and guiding them while they are following these instructions. The teachers who quoted this role often mentioned it is important for learners to be actively involved in this process, with the teacher being there to prepare learners for learning activities and providing them with support when they get stuck.

"From my experience, the teacher is there to give guidance, and give directions and instructions to a learner. Most of the works in the classroom should be done by learners and not teachers like the way it used to be in the past".

"A teacher is supposed to give guidance… What do I mean here by giving guidance? There is a preparation stage, and, in every activity, the teacher is supposed to introduce such activity, to give instructions, also to give guidelines so that the learners can follow – because it is also about learner participation."

Regularly, teachers saw their role as a supervisory role, someone who needs to make sure children are disciplined and pay attention. This role is however never considered the only teaching role; whenever teachers mentioned this, they also mentioned other roles pertaining to them.

"From my experience I have as a teacher, the role of a teacher in a classroom is to arrange the environment for learning, to support the learners whenever there is a need to, and to provide discipline to the learners."

"The role of the teacher is to equip knowledge to the children and make sure they pay attention in class."

In addition to that, teachers would also regularly mention that the teacher has the responsibility to make sure the learning environment is conducive. They referred to the larger classroom environment being clean and organized, and to the provision of proper teaching and learning materials.

"The role of the teacher in a classroom setting, number 1 is to make sure that the furniture and setup of the group is well-centred in the classroom. That is important for easy movement, also to allow a free collaborative learning. And also the teacher should be able to be distributing learning and teaching materials on time, because of a very good setting and it will also help the teacher to maintain order for smooth learning for the learners."

Besides these more frequent descriptions of teacher roles, there were some other roles which are interesting to mention. Incidental mentions were the teacher as mentor or counsellor of leaners, the teacher as a role model, someone who prepares learners to be good citizens, and the importance to assure quality education.

When looking at the answers provided by teachers, attention was paid to the level of agency that was assigned to learners in their answers. In roughly 17% of the interviews, teachers assigned shared agency between teachers and learners in the classroom. This is for example shown in the earlier two excerpts on the teacher as a facilitator role, which both stressed the importance of learner involvement.

Incidentally, answers would stress the importance of leaving no learner behind in the classroom. These responses would state that the teacher has the responsibility to assure that all learners can participate and learn.

"The teacher basically is the manager – the teacher is supposed to oversee all the running of programs in the class to ensure that each and every learner gets the best out of learning regardless of their challenges as you know that different learners have different abilities. The teacher is supposed to make sure that they can bring the best out of whoever is being taught in the classroom."

In addition, a small number of answers would look at the holistic development of the child and seeing the teacher role as more than just impacting learning.

"The role of a teacher in a classroom setting is to nurture those children into good citizens. Make them become responsible people."

All in all, answers of teachers are focused on learning in general, and less on holistic development of a diverse set of learners in the classroom. While imparting knowledge is still considered one of the most important roles of teachers, answers also show that teachers find active involvement of the learner important. Many teachers refer to themselves as facilitators, thereby assigning a more active role to learners than simply paying attention. This is further corroborated by some teachers specifically referring to learners' agency in the learning process. Active involvement is a key component of the Catch-Up program and seeing teacher answers reflecting this when being asked about their role, is encouraging.

For stakeholders, imparting of knowledge was the most often cited role for a teacher. In addition to that, ensuring a conducive learning environment was also often quoted. Other roles that they mentioned which were congruent with what as mentioned in teacher interviews was the teacher as a facilitator role, and the teacher as a mentor role. Similar to the teacher answers, there was some attention for the shared agency between teachers and learners in the classroom setting. Interestingly, in these interviews ideas of holistic development and leaving no-one behind were more frequently mentioned than by teachers. The excerpt of an interview below shows how this stakeholder understood both things to be core to the teacher role:

"So the role of a teacher actually begins where children are struggling. Not those children who are already reading and are doing very well. You can't see the role of a teacher working so well. So in short, the role of a teacher is to shape and reshape the direction of learners' thoughts or thinking, so that the learner can use that which he acquires in the classroom situation to use it outside the classroom situation, it must not only be academics, but a lifelong kind of thing. A teacher's role doesn't end in class- the knowledge that the learners acquire needs to be utilised in their communities so that communities can be appreciated."

Follow up 1 teacher: What is the teacher's role in classroom management?

This question is similar to the previous one, however it switches from the general classroom setting to specifically managing the classroom. This similarity might explain why in half of the interviews this question was unfortunately skipped. This should be considered when interpreting the comparisons that are made between answers to this question and the previous question.

In general, similar answers were provided as for the previous question. However, for classroom management the two most important roles of the teachers were supervising and disciplining learners and creating a conducive learning environment. Again, with a conducive learning environment

teachers meant both creating the right physical environment as well as providing quality teaching and learning materials. After these two roles, teachers also regularly described roles which did not come up when defining the teacher role in the general classroom setting. The first one is the teacher assuring clear lines of communication within the classroom. The following excerpt also describes what the result is when the teacher takes up this responsibility:

"The role of the teacher in classroom management is… and communication. So the role of the teacher is to ensure the class is well managed for the benefit of the learners so that even a learner who is a slow learner is able to grasp something from the interactions, as they interact with each other."

In addition to the communication role, teachers also regularly mentioned that teachers are responsible for time management and assigning leaners work:

"Normally what we do for them to be managed is we give them work to keep them busy."

"Yes, and then in classroom management you need to have time in class, when we are teaching in class, we follow the minutes assigned on the timetable."

Besides these new roles, teachers also regularly mentioned imparting knowledge as a responsibility of the teacher, as well as acting as a role model. In general, it looks like for classroom management teachers are more focused on teacher roles that provide structure and organisation to the classroom, as to assure that learning can take place according to schedule.

While in these answers there is less focus on learners' agency than when defining the teacher role in the general classroom setting, there is a slight increase (percentage wise) in answers making mention of student's holistic development and leaving no student behind. However, these changes are quite small and hard to interpret with multiple interviewees not having answered this question.

As for stakeholders, their answers are not clustered around specific teacher roles. In general, they refer to the same roles as mentioned by teachers, or as mentioned in the previous question on the classroom setting. Some interviews mentioned roles which were not yet mentioned before. The interview excerpt below shows how one stakeholder views that classroom management should take place:

"…So maybe this is where we have agreed, the learners and the teacher will agree on how will we manage ourselves? So what will be our classroom rules? So things that they agree on, so this is how we are going to manage ourselves, so that we have order in the classroom so the rules will come in but then the teacher should take care of maybe the mental wellbeing of learners, their emotional wellbeing. So this will be done maybe through the daily greetings; is anybody sick? Is anybody … and how's home?"

This excerpt shows how this stakeholder views collaborative rule setting between teacher and learners as essential, as well as the need for the teacher to look after learners' mental well-being, rather than just being focused on their academic achievements. Later in the interview, this stakeholder mentioned:

"You'll see them (teachers) *posting pictures of how they are greeting their learners and so everybody is taking it upon themselves and say okay I'll find my own way of you know having that kind of relationship with my learners. And from a few teachers that I've coached, they also bring out to say I'm also working on how I'm relating with my learners, so you begin to see that teachers are conscious about how the learners are in terms of how are you? Or how is this person before I can start teaching…"*

In contradiction to the interviews with teachers, interviews with stakeholders do not show a clear pattern as to what is understood to be the role of the teacher in classroom management. Interestingly, they all hold different ideas. However, both teachers and stakeholders do not see classroom management as a process of strictly supervising and disciplining learners but understand the importance of providing structured and conducive environments, with clear lines of communication, as well as occasionally recognizing how their practices can influence holistic development of the student and assuring no-one is left behind.

Follow up 2 teacher: What do you think is the most effective way to teach children?

Almost half of all interviews mentioned that activities should be learner centred. Multiple beneficial results were mentioned of actively involving learners in the teaching process, of which some are described in the excerpts below:

"The effective way to teach children mostly is when learners find answers on their own, it's a way for them not to forget. Because that method helps the learners to find answers than you the teacher finding the answers for them."

"The lesson should be involving both learners and teachers, as much as it is involving with play. Like the way it is with Catch-Up: the teacher is involved, the learners are involved. No one is left behind so that one is answering an effective type of learning. It gives good understanding to learners."

"Yes, the learners are involved and participating in the activities. As a teacher you are just guiding, but as learners they are to participate fully. That is what we call learner-centred, where every learner is involved. I think this is effective because it encourages even slow learners to participate."

These teachers mentioned how learner centred teaching methods help them to engage all learners and promote understanding and retention of knowledge. After learner centred methods, playful method was the most often mentioned by teachers. This includes a wide array of playful activities: games, toys, role-play, the use of stories, singing, and dancing. For playful activities, teachers mentioned how they bring joy to learners, promote all learners to be engaged, and aid learner understanding:

"I think by using play, role play, songs, where children will participate and enjoy while they are learning."

"The most effective way of teaching children is to engage them into role play. A teaching that involves a certain kind of play, which makes them understand better what you as a teacher are teaching them."

After learner centred and playful activities there were 3 other categories of effective teaching that were mentioned equally often. These were making things practical, using learning and teaching materials and the use of group work or another way of peer-to-peer learning. As for the practical learning activities, teachers mentioned that it helps learners to understand as well as promotes retention of knowledge. One clear example of how to make activities practical, is described in the excerpt below:

"It's to let them do what you are telling them to do, like there's a practical maybe in science then you want to experiment on liquids and solids. Or the things that float or sink because they are in lower primary. So you tell them to bring whatever is used such as water, stone, plunk or something like rubber. Then you tell them can you put water in a dish and again water in another dish, then you tell them to put a stone and another a rubber thing or plunk, then ask them whatever they have observed from the experiment."

This teacher also shows how natural materials occurring in learners' every day environment can be used in these practical activities, which is especially important in resource poor settings within which some of these schools are located.

Teachers see the use of learning and teaching materials to guide and extend the learning process, as can be seen in the excerpts below:

"Yeah, there are certain strategies that can be used where we give teaching and learning materials to learners, then from there they come with their own ideas, then from there you as a teacher start expanding on those ideas. The learners also start researching on the given ideas."

"…You are there just to help them (learners) find the answers and you should have a lot of teaching aids because those teaching aids are what assist learners to find answers or to read."

Besides learning and teaching materials being mentioned as a direct response to the question about effective teaching, in another interview the use of appropriate and engaging learning materials was seen to promote classroom management when teachers are absent. The teacher explained that when the classroom includes learning and teaching materials, learners can keep learning even when the teacher is not around. In a stakeholder interview, the use of materials in combination with storytelling was even seen as an effective way to keep learners coming to school every day, because they want to know how the story continues. In summary, the use of teaching and learning materials is one of the key effective teaching methods teachers have at their disposal. Importantly, one teacher who also mentioned the effectiveness of teaching and learning materials, stressed that these were not provided

sufficiently. An excerpt of the dialogue between the interviewer and the respondent illustrates which challenges this teacher faces:

Respondent: "There are no materials for Catch-Up and to teach effectively they should provide more materials for learners" Interviewer: "What do you mean by the term they?" Respondent: "It is the Ministry, the school provides some materials, but the government should provide some materials and other things so that the program can continue, for example, teachers carry their own materials to teach Catch Up." Interviewer: "In Catch Up learners use local resource materials. Is it possible for learners to be engaged to provide materials such as lids from container and stones when teaching numeracy?" Respondent: "We normally do that, but learners come from different backgrounds, for instance, some buy cooking oil in buckets hence they do not have lids."

This excerpt shows it can be challenging for teachers to come up with materials that are to everyone learner's disposal, adding an additional burden to teachers to be resourceful in these situations. Not all teachers might have the bandwidth or the motivation to go that extra mile.

Lastly, another frequently mentioned effective teaching method was the use of group activities. Most of the answers referring to group work, mention that it allows learners to learn from each other, especially in situations in which they might not be understanding what the teacher is explaining:

"I think from observation group work really helps them, because if they can't get it from the teacher, they'll be also learning from their friends. Some learners it's much easier for them to get it from the friends than to the teacher. So I feel group work for me is good, is the effective one."

"Through collaborative where children give each other ideas, where they are in groups – they share ideas. First the teacher has to offload information, then forms groups for learners to share ideas. After sharing ideas, they present. You'll find by doing these steps learners will grasp certain concepts produced in every lesson… Kids they learn best when they are able to share ideas with each other, because there they don't even feel shy. They are highly engaged, cheerful and highly connected, because they talk to each other and play together."

Thus, according to teachers, group work can help shy and slow learners to be up to speed with learning, and in addition it is an activity which learners enjoy.

Besides looking at what teachers consider the most effective way to teach, we also looked at what teachers mentioned are the results of using these ways of teaching. Even though teachers were not asked this specifically, excerpts above show that teachers often did explain why they considered certain teaching methods to be effective. This gives us a better understanding of what "effective teaching" means to teachers, and how they understand the learning process. The two things that were mentioned the most were methods that were able to promote understanding and assured that no one was left behind. The last category focused most often on methods that were able to involve shy learners, who would otherwise not participate. In addition, it also sometimes referred to slow learners, who would not be able to keep up if the effective methods were not used.

After these two most important outcomes of effective teaching methods, 3 other outcomes were mentioned in roughly 10% of the interviews: peer-to-peer learning and support, active participation, and enjoyment of learners. The last excerpt describing the outcomes of group work nicely captured these 3 categories. In addition, the excerpt below shows how the use of playful methodologies promotes learner enjoyment and peer-to-peer learning:

"…Involve them (learners) *maybe in fruitful playing like what Catch-Up does, because learners enjoy learning when they are playing. That's why it's easy for a child to learn from a friend, unlike with the teacher who would just strictly be focusing on what they want to do, but with their friends they'll be learning as well as playing."*

Lastly, two outcomes that were mentioned more than once was assuring that learners would not forget what they learned, and learners being able to learn faster than they normally would.

Stakeholders had similar responses to teachers. They also mentioned the importance of learner centred and playful methodologies, as well as using learning and teaching materials and group work. A new observation in the stakeholder interviews was the importance of using the learners' language of play to promote their learning:

"…also the use of familiar language because communication is very important in the learning process. Children learn best in their language of Play. In Luangwa, the dominant language is CiKunda but the children are examined in CiNyanja because of the zone, so when we're monitoring different sections you see the teacher begins in CiKunda when teaching the children and then brings it to CiNyanja, so this really excites children to learn because they are first learning in a language that they are familiar with, a language that they use when they are playing and this way they are able to understand what is being taught. In some cases, you find children all the way up to grade seven, grade nine who are still not able to understand the official language of instruction which is English so teaching them in a base of a language that they are familiar with is very helpful to their learning process. So at times, even when a teacher is teaching in English, they would revert back to the local language just to explain some things and give the children clarity. So this way the teacher accommodates children who may be slow or who may not be fully competent in using the official language."

In Zambia, Catch-Up is only taught in the home language, which is most often the same as the language of play for children (although not always).

In terms of outcomes of effective teaching methodologies, stakeholders also mentioned that their use promotes learners' understanding. They did not mention the other outcomes that teachers came up with. In addition to promoting learners' understanding, one stakeholder mentioned that effective methods promote the acquisition of skills as well as encourage children to try out things practiced in class for themselves:

"When they learn through play, I see children in the community trying to do exactly what they did in class, even when they're playing in their own settings at home. So they role play the same things that they've been learning in class. And this helps them acquire a lot of necessary skills."

In summary, it is positive that teachers and stakeholders are able to mention a diverse range of activities that can engage learners, different from the traditional copying from the board and memorisation teaching methods. Teachers recognize the importance of active learning and playful activities, and they can clearly articulate their beneficial outcomes on a learner level. While this is not a quantitative analysis, it is encouraging to see that fostering understanding was mentioned more often than not forgetting. In addition, teachers equating leaving no-one behind with effective teaching is in line with what the Catch-Up method wants to achieve. Catch-Up is specifically focused on ensuring that learners from different levels all receive teaching that is appropriate for them. The focus on group work as effective teaching, and the often-cited positive effects of peer-to-peer learning also shows that teachers recognize the agency that learners can take in not only their own learning, but also the learning of their peers. Recognising learners' agency and providing them with agency when appropriate is a key component of the Learning through Play approach (Zosh et al., 2017). All in all, when it comes to effective teaching, answers provided in these interviews are in clear alignment with the Catch-Up approach.

Follow up 3 teacher: What activities can the teacher engage his/her learners in to enhance their participation in class?

First, this question was asked in general, followed by a set of 7 probes meant to ask about activities aligning with each of the 7 components of the 7 Cs framework.[28] We will first have a look at what teachers and stakeholders mention without receiving targeted probes, to identify which activities are most salient to them when they think about promoting learner participation.

The most popular activity was by far group or pair work, which was mentioned in half of the interviews with teachers. This could be group work with or without a competitive note. Teachers again mentioned how group work promotes shy and slow learners to participate, learn, and express themselves. Also, teachers mention again that leaners can learn from each other. The benefit of the competitive element is nicely illustrated in the excerpt below:

28 The following probes were used: Concrete: *"activities that are engaging"*, Captivating: *"activities that attract learner's interest"*, Connected: *"activities that relate to a learner's (already existing) interests"*, Challenging: *"activities that appeal to a learner's creativity and invoke imagination"*, Creative: *"activities that allow learner's freedom to express themselves"*, Collaborative: *"activities that encourage learners to work with others"*, Cheerful: *"activities that bring joy to the children"*.

"And also you can do it in the form of a competition whereby you divide learners in groups and from there they'll be giving words maybe in literacy, they mention these words they write, by doing so, they'll be comparing which group is doing better, them not knowing that they are learning. This will enhance their betterment. Children will fight by all means to become victorious. By doing such interesting activities in a funny way, whereby children are enjoying, they are not dozing, learners will be able to move from one stage to another without even finding challenges."

After group activities, class level activities like discussions, presentations and debates, and role play were mentioned in roughly 20% of the interviews as engaging activities. For class activities, teachers see it as a way to get everybody involved in the activity. Role play was usually mentioned without explaining why they thought this method was engaging. The third most quoted engaging activity was using practical activities, meaning learners get the opportunity to directly engage with the materials they are learning about, or when abstract concepts are put into concrete activities with materials they can touch and/or manipulate. The excerpt below clearly explains how this happens in the classroom, and which benefits this has:

"And you'll find that it's hands on for the learners. For example, if we are to do numerals using sticks, you'll find that those sticks will be given to the learners, they'll do it themselves and they'll lift them, count, and put them on one side on their own. If they are to draw boxes as they shape, you'll find that they are doing it on their own, they are starting it on their own, drawing on the floor or on the ground. When we are reading, we give them reading cards, you'll still find that those cards are being held by the learners themselves. And through that they are drawing interest into it. You are also encouraging them to be part of the lesson. Other than that, it's like you are motivating them and making them feel like part of the class/lesson other than where you are just talking like you are lecturing and you are doing everything as a teacher."

Making activities learner centred as well as using playful activities were also regularly quoted. Playful activities included games, toys, and sport games. It is important to note that in this question, role play and storytelling were considered separate categories from playful activities, while in the previous question these categories were included in playful activities. This was done because in this question we asked about specific activities, while in the previous question we asked about overall teaching methods, which resulted into broader categories in the former question.

In the interviews, one teacher explained how the use of playful activities makes learners forget that they are learning, which is quite similar to what was described in the previous excerpt about the use of competitive group activities. Besides these more often quoted activities, other engaging activities mentioned were the use of exercises or homework, field and outdoor activities, storytelling, specific mathematical activities as taught by Catch Up (e.g. Number Circle Activity, Mind Maps, Flash Cards), and arts and crafts.

It is interesting to see that there is such a preference for group work as an engaging activity. It might be that teachers through Catch-Up have discovered how the use of groups can be beneficial in terms of engaging learners. For the rest, all activities mentioned show a relatively high level of learner involvement, with some teachers specifically mentioning how these methods are more effective that the traditional teacher instruction methods. There is also a strong reliance on playful activities. Even though we split up categories here to look at role play and storytelling as individual sets of activities, when merging these we see that more than one third of the interviews recognize playful activities as engaging.

Stakeholders have very similar responses to teachers, but their responses are more high level, in the sense that they refer to characteristics that engaging activities should have, more than naming specific activities like teachers would do. They often appear to be very aware of the 7 Cs framework, which is reflected in their language use. One stakeholder also points out that relationship building between a teacher and learners is key for learners to be engaged in the learning process. The excerpt below shows what the teacher means with this, and is also an illustration of stakeholders' understanding of the 7 Cs framework:

"Another thing is maybe from the social arrangement because you know, the teacher also needs to make sure that he takes care of the emotions of the children. They need to be free. The language has to be appropriate. The children need

to be respected. They need to be praised for their effort and may need to be given enough time to share the experiences. And all the time when the teacher is teaching, it must move from known to the unknown. So a teacher just can't come in and start teaching from where the children have no idea. The teacher needs to begin from where the children are, and then move up to teaching them about what they do not know."

Importantly, teacher answers do not necessarily reflect the phrasing in the 7 Cs framework, however, their explanations do show that they put the 7 Cs in practice, which can be seen in the excerpts above. This is probably simply a reflection of stakeholders engaging with the actual policies, while teachers are more focused on implementing these guidelines.

After specifically probing for engaging activities aligning with each of the 7 components of the 7 Cs framework, group work activities, practical activities, class level activities and role play all continued to be among the most often mentioned. Playful activities and storytelling were mentioned more frequently when specific probes were used and were now among the most quoted activities. In addition, teachers more frequently mentioned very specific activities which appeared to come straight from Catch-Up guides (e.g. Number Circle Activity, Mind Maps, Flash Cards). The use of exercises, reading, arts and crafts, and field work and outdoor teaching were also more frequently mentioned, however they remained among the least mentioned activities. The use of multiple probes gave the teachers an opportunity to reflect on the wide array of activities they have at their disposal, which explains why most of the activities were mentioned more frequently after probing than before. It is encouraging to see that there is such a variety of activities used not only between teachers but also by individual teachers. Even though having to respond to one main question and 7 additional probes about engaging teaching methods, most teachers responding to all probes were able to keep coming up with new methods they used to engage their learners.

As for stakeholders, after being probed about alignment with the 7 Cs framework, three new insights came up. One was that most of the interviewees mentioned the importance of using appropriate teaching and learning materials when this did not come up before probing. In one interview the stakeholder mentioned how the use of wall posters and pictures, in combination with storytelling, motivated him to return to school every day:

"For example, I still remember when I was in Grade 4 way back in 1974, my female teacher Mrs Daka bringing what we call the wall posters/pictures. What can you see on this picture? She asked. Those pictures made me go to school and like school more and more. And we were discussing with my colleagues, "Look at Mulenga, he has kicked the green ball. Why didn't the teacher run away from the green ball when it was coming But Mulenga kicked the boy and it hit the teacher. Now again, you see that Kalumbi and Welengani fell on the bicycle." So tomorrow I want to go back to school to see what wall picture the teacher has brought. So such salient pictures bring in a lot of joy in the hearts of learners – the hearts of learners are actually captured. Yeah, because they want to see what is next tomorrow."

Another stakeholder described how important it is to make sure to use materials that are appropriate:

"But also then the contextual activities that may be things that are similar in their environment. So take for instance like I've noticed like puzzles, in our Zambian schools especially our public schools, not all children are really exposed to puzzles in their homes. So whilst a puzzle would be a very engaging activity, sometimes especially in the beginning it's strange to them so you don't see them have an interest until they get used to it. So the teacher also needs to think about what are these children engaged in in their contexts, so we can use that kind of game."

The explanation of the stakeholder shows a clear understanding of the Connected component of the 7 Cs, which states that methods and materials used should connect with the social environment of learners, as well as their existing knowledge and interests.

Another thing that came up in most of the interviews was allowing learners to participate in free play. This was about learners having the ability to choose what type of activities they want to engage in, even if activities are done at a class level. One stakeholder mentioned that teachers can let learners

choose democratically, but should also be observant when learners get to choose their own activities, to see what type of activities they choose to engage in. Another stakeholder described how limitations put on learners' play, as well as only allowing learners who have ideas congruent with the curriculum to share in class, impedes their creativity and imagination:

"Because as teachers, we like saying no. This is not the answer and children are demoralized. But when a child expresses himself or herself to a near answer, or different expression from the answer you want, call it the answer. But you say no, there is even a marking key. If you don't mention all these four items, then you have failed. That doesn't lead to creativity, innovations and imaginations because a learner will say even if I write Doc Mooya will mark me wrong."

"…Whereas the teacher does not need to draw a line to say whatever you want to think about, I want you to end here. No, let it be as open as it could be, and this is how we see different phones, different laptops being developed, different types of house fans, different backpacks, different shoes. Because those children have got that kind of imagination. What if I make a shoe like this? Will it sell? What if I do like this? They have taken that creativity and now they are innovating on it, coming up with so many different designs in many quarters. So a teacher needs to have that liberty life to be given to the learners."

"Even at your home Doc, buy a toy car to your boy child, buy a Doll to a baby girl. You will find that they have dismantled it. We cannot say what we buy them is a wrong thing. No, they are three plus. We want them to be playing with such a toy, but to them they have gone beyond, they want to see what is inside, which we do not allow. Even as parents, this one is stupid. Each time I buy a toy, it does not take long, in just two days it is destroyed. But we are forgetting the imagination part."

Another stakeholder echoed the need to have all learners share their ideas, regardless of whether they are wrong or right. Teachers should not just be focused on correcting whatever is not in the curriculum, but allow learners to come up with alternative ideas, and assure that the way answers are corrected, does not discourage learners from sharing again in the future. These answers of stakeholders show a clear understanding of the 7 Cs framework, as well as an understanding of what learning through play means, how it can be implemented, and what benefits it has for learners.

Main question learner: From your experience, what do you think is the role of the learner in a classroom setting?

Most of the answers show either a passive role for learners, or an active role, but only in reaction to what the teacher is doing. The most often mentioned responsibility of learners is to participate fully in classroom activities, without further elaborating on how that participation looks like. However, these answers show little agency of learners. They simply have to take part in whatever activities are initiated by the teacher. Other activities that were quoted in roughly one third of the interviews was that learners should pay attention in class, follow the instructions and rules set by the teacher, and should learn in class, understand what they are being taught and perform well. All these activities show a very passive role of the learner, listening to and following whatever the teacher is saying and doing. The excerpts below summarise these ideas:

"The learner is there to do what they have been instructed by the teacher and to pay attention, because without paying attention, the learner cannot do what the teacher has instructed."

"The role of the pupil is to listen, pay attention, participate and be a responsible learner who will fit society."

"The role of the pupil is to perform and behave all the time and put in place everything they learn in order to perform."

"The learner has to participate by answering the questions and obeying the rules."

In addition to the learner responsibilities above, other roles that were mentioned by teachers regularly was to behave well (which can be seen in the excerpts above), to take notes, and to be present and on time.

There were some teachers who mentioned more active responsibilities for learners, but even though they required more involvement from the learner, around one third of the time these responsibilities were still in relation to whatever the teacher was doing. Teachers mentioned that learners should answer questions in class. Other learner tasks which were mentioned incidentally, was that they should help the teacher, clean the classroom, or encourage other learners to fully participate.

Most teachers did not explain why these roles were assigned to learners, but one teacher explained how they go about setting rules and why it is important:

"Us who teach the lower ones create what we call 'classroom libraries' and then there's a certain period that is set specifically for the library. Usually, we give them free time to go and pick reading materials on their own, maybe they can be in pairs or groups – they do the reading. A learner is supposed to do that his/her own unlike when they are being forced. So that is the role of a learner. A child is supposed to know this time I'm supposed to be in the library, this time I'm supposed to be doing this so that at the end of the day this child becomes a responsible child and that each and every time he/she follows the classroom routine."

From this explanation, the creation and following of rules is seen as necessary for this teacher to promote learners becoming a responsible individual, who can follow rules without having to be reminded of them. This explanation refers to a certain level of agency for the child, since the teacher is not continuously enforcing these rules. This is reflective of preparing children to operate in society with laws, norms, and values, in which there is not someone constantly reminding you of what you should do in any given situation.

Besides the more passive roles of learners outlined above, roughly one third of the interviews mentioned more active roles of learners, with a higher level of agency. Teachers regularly mentioned that learners should do work on their own and should research the topics that they are taught in class. The excerpt below mentions how this is especially important in the African context:

"For a teacher I said she's a guide. Now if the teacher is a guide, then a learner has to research, to participate, even to answer most of the questions given to them by the teacher. So their work is mainly to research, discover and learn more. Kaili the reason I say this is the issue of spoon-feeding children killed most of us African countries. Our African children will just go in class, give them 1+1=2 mwanvela? Yes. 2+2=4 mwamvela? Yes. But if you give a child to say what makes this stick and this stick 2? The child will answer: "Kaily, we are dealing with addition. When I add these sticks together will give me 2." You'll find that the child would have discovered or researched the question herself and come up with the answer, that the answer is 2. This is where it is coming from."

In addition to that, teachers state that learners should ask questions to the teacher and share and discuss their ideas with the teacher and the other learners, because this is beneficial for the learning process, as explained by one teacher:

"A learner is supposed to be attentive, co-operate, participate. Give information about what he/she knows because these children are bankers of a lot of information. Yes, they know a lot and that is why we should just be brief simple, and then will be able to get a lot from them and then work on that. If they give us little information, we now expand from there because they also know something."

The teacher not only recognizes and values learners' knowledge, but also shows how they make use of the 7 Cs framework, by building up on existing knowledge that learners have already.

Stakeholders have very similar responses to what teachers have mentioned above, but when it comes to the following of rules, there is more often a focus on how this is a collaborative process between teachers and learners, as the excerpts below show:

"Next is to be following the instructions that are coming from the teacher or in this case, if they are in groups from the leader."

"I think the learner has a responsibility. Their role is to maybe try to follow what they've agreed as a class. I think they owe it to the class to be a member of that class, to be a member of a team okay and so for you to be a member of that class, there are certain behaviours that you need to portray. There are certain rules, certain restrictions that you put, you need to say okay I cannot go this far, how do we speak to each other here. How do we respond when that person is sick or when they've failed in that class. They haven't given a correct answer so how do we surround ourselves? So those are some of the rules they have to follow. I don't want to use big words like empathy for the learners. But then it's … if our friend has not done well, so what do we do in this class? Yeah. So it's bordering around our rules. So the learner has the responsibility to obey the rules set for themselves."

The last answer shows that rules set in a class should go beyond the learning process, but also look at learners' well-being. Importantly, it mentions how learners have a responsibility to look after each other.

In general, we thus see that teachers mostly see the role of learners as relatively passive, following and reacting to whatever a teacher does and says in class. This is not that surprising. Given the low grades in which Catch Up was implemented, it does make sense that the level of agency of learners is not that high. In addition, Catch-Up is not specifically focused on promoting agency, with the teacher being the one to select the sequence of activities, with a strong focus on planning and time on tasks. It is also important to note that learners' agency in the learning process is a concept that is understood differently in different cultures. A study from Nduku, Staskowicz, & Stern (2002) showed that more subtle markers of teachers' support for learner agency exist in cultures where support for child agency is not the norm. Nevertheless, there are still a reasonable number of teachers and stakeholders who do foresee more active roles for learners and are able to recognize the benefits these active roles have compared to more traditional learner roles. Many of them show the ability to link this back to what they have been taught under the 7 Cs framework.

Follow up 1 learner: What is the learner's role in classroom management?

This question is similar to the previous one, however it switches from the general classroom setting to specifically managing the classroom. Like what happened for the teacher questions, this similarity unfortunately caused this question to be skipped in 8 of the interviews. Again, this should be considered when interpreting the comparisons that are made between answers to this question and the previous question.

The answers provided by teachers remained focused on more passive tasks of learners. In terms of classroom management, more than half of the teachers expected learners to follow instructions and rules.

"The role of the learner is to do what the teacher… or to follow instructions that the teacher has given."

While the majority of teachers do not elaborate beyond stating that learners should follow instructions and rules, a few teachers explain why rule setting is important, and that learners should be contributors to and enforcers of these rules:

"A noisy class – it is their responsibility again to see and to know why they are in class. You know a noisy class is always distracting to those who want to learn so you ask them to come up with rules to manage the classroom. By doing

so, it will help them whenever they want to go off their way, they'll look at the rules that they themselves have put up. So it will help them in a way."

After following instructions and rules, other things that were mentioned regularly were quite similar to the previous question, including that learners should learn and understand what is taught in class, pay attention, know why they are in class, and behave well. One teacher describes why it is important learners should understand why they are in class:

"The learner should know why they are in class. When a learner knows why they are in class and the importance of why they are in class, then class management will be easy. Because there are some learners who know why they are in class and those who do not know why they are in class. You will find that they are the ones who will be telling their friends saying "You, can you keep quiet, can you keep quiet", such things. And you will find that they are the ones even helping the teacher in classroom management."

Also participating fully was still mentioned frequently. The excerpt below describes why this is an important role for learners:

"Now if this child doesn't participate in writing, how will the teacher know whether the child is making progress or not? So he/she has to do their part when it comes to writing. If the child isn't clear in class, the teacher has talked and maybe that child hasn't understood a certain concept, a child has that duty telling the teacher to repeat the topic or that segment so the teacher can use a different method or find a better way of helping that child."

The previous two excerpts show that while many teacher answers attribute a seemingly passive role to learners, when teachers elaborate on these concepts, these answers sometimes show that there is a certain level of involvement expected from the learner, beyond strictly following teachers' initiatives. This shows that some teachers might have a less passive image of learners than some of these answers might suggest. In future research, it would be interesting to further unravel this by asking targeted follow up questions that would allow for a better understanding of how teachers see learners' involvement and level of agency in activities that seem more passive at face value. Importantly, this should be integrated with the literature on the understanding of the concept of learner agency in learning through play in other cultures (Nduku, Staskowicz, & Stern, 2002).

Lastly, there was again a reasonable portion of answers that attributed a higher level of agency to the learners. Some of these were still focused on learners responding to teacher requests, like answering questions and cleaning the classroom. However, there were also activities mentioned with a high level of learner agency. Interestingly, some activities were identified here that did not come up during the learners' role in the general classroom setting. While the sharing of ideas in class was mentioned in the previous question as well, this time teachers also named organizing time, reading, and interacting with other learners as important activities. Organizing time was also a role that came up for the teacher's role in classroom management, showing that time management is considered important for both learners and teachers alike. When talking about interacting with others, teachers pointed out that children learn better when they get to interact with their peers, which resonates with what teachers discussed about effective teaching methodologies.

"It is important for each and every child to interact well with fellow learners, because learning, you'll agree with me, is not just gotten from the teacher's point of view. In fact, children learn best amongst themselves. So interacting with other learners is very important… There is even other information that might not come from the teacher, which they can share among themselves. There are those cut-crossing issues that children are able to share among themselves."

Stakeholders had similar answers to teachers. One new thing they explicitly mentioned is that learners need to provide leadership to their peers.

"The first one is that of leaders because when they have groups, there are some group leaders. And it's an opportunity for them to learn how to lead others and that leadership will change from time to time depending on their activities. And depending on their understanding and when there's exchange of leadership, it fosters respect for one another. But this of

course is within the setting of being given instructions and having the children carry out the instructions that the teacher has given."

This shows how the use of groups is not limited to a teaching methodology but is also used to give learners active responsibilities in terms of classroom management. This answer also shows again how a seemingly passive learner role (following instructions of the teacher), is actually understood by the respondent to be more active than one might expect at face value.

Follow up 2 learner: What strategies do you think children use (best) to make sense of the information they receive in class?

Teachers mentioned a wide variety of strategies in response to this question, and the frequency with which these strategies were mentioned was relatively evenly distributed. More than half of the teachers focused on strategies that were initiated by them, while the rest of the interviews mentioned strategies that were initiated by learners themselves. This shows that teachers have differing views into how much direction teachers should provide to learners making sense of information.

Strategies that were mentioned to be initiated by both teachers and learners was role play and practicing and trying out what they have learned. Strategies initiated by teachers to help learners make sense of the information they receive included group work, songs and rhymes, exercises and homework, taking notes, and asking questions as well as soliciting feedback from the teacher. These strategies are similar to the effective teaching methodologies mentioned by teachers before. In most of these interviews, teachers do not elaborate on how the strategies used relate to the learning process, which makes it difficult to understand how these teachers view this process. For the small number of answers that do elaborate, teacher answers relate to the Cheerful and Collaborative components of the 7 Cs framework:

"Children enjoy when you put certain concepts in rhymes and songs. They begin to master and also repeating more challenging concepts. Let them do it again and again and again in group work, so that certain concepts can be grasped."

"They like for example you as teacher to explain to them, then they can do role play, they can come and do what you do. Because most the time they enjoy that."

Strategies initiated by learners included play, interacting with what they learn, engaging with material independently from the teacher, discussing what they have learned with other learners, and relating and applying new information to real life. The excerpt below describes how a teacher sees that learners apply and engage with information even outside of the classroom:

"The strategies they use is maybe coming back to the teacher and say what you taught us yesterday or what we learnt, I went home and did ABCD…I'll give an example, during the times of Covid, we used to tell learners to say "You need to follow the golden rules". Masking up, social distancing, and all those. So when the learners are told that, when they go home they will try to put that into practice what they have been taught. And they will also try to disseminate that information to their parents or other people in the community. In mathematics, if maybe you had given them a problem to solve, they'll go home and maybe those other gifted children will try to look around other books and find a similar problem. They will try to solve and bring it to you and say "Oh teacher, I found this and I thought I should try it on my own and bring it to you so that you can counter-check."."

Another teacher described why it is needed for learners to interact with the things they learn about:

"They have to apply it because sometimes memorized information can be forgotten. They need to apply it because Catch-Up for them is all about teaching things, being creative and being motivated for them to do the work. They make sure that they touch what they have seen and they won't forget about it the next day. It is better than them memorizing when they won't even see it. Because after some time they will forget about it. When they do their work on their own, you find they will be unable to forget."

Both teacher and learner-initiated strategies suggested map with how the 7 Cs describe children's learning process to take place: through concrete activities that relate to what children already know, making use of peer-to-peer learning and their senses, and participating in joyful activities as well as trying out new things through trial and error. While teacher answers are not elaborate enough to get

a grasp of how teachers understand learners' learning process, when they do, most of the time teachers seem to be focused on surface level learning: learners not forgetting what they are taught. The except above is a clear example of that. However, some answers do show deeper levels of learning. They go beyond just memorising and talk about learners applying what they have learned to different situations, new problems, and outside of the classroom. The previous excerpt is an example of this, describing how learners actively find similar mathematical problems without the involvement of the teacher and are able to solve them themselves.

Asking teachers more specifically about how they define learning and what processes are underlying would give us a better understanding of where on the spectrum of surface to deep learning teachers' definitions lie. The current study only indicates that teachers differ in their definitions, and that the first instinct of teachers when thinking about learners making sense of information might more often be focused on not forgetting rather than deeper levels of understanding.

Stakeholder answers are very similar to teacher answers, however, they more frequently elaborate on how the suggested strategies relate to the learning process. From these more detailed descriptions it also shows that they see learning as more than surface level. The excerpt below gives a detailed description of how one stakeholder views the learning process:

"Because if they're able to learn from the teacher but also from one another, it makes it interesting for the children because these children come from different backgrounds, one may be coming from mayadi, one can be coming from a compound or one can be coming from village setup, and when they come together, they are able to share different experiences and make sense of what the teacher would instruct them to learn when it is related to what they know and what they have experienced. Some of the children especially for example, Zambia being a Christian nation, some of them have different experiences in terms of spiritual exposure or teachings that the teacher may not be aware of, so once they start learning from class, the children then realise okay, so there are people who know some different things from what I know and then this helps them to start relating. So this new knowledge the children are exposed to also gives them the curiosity to research and to ask questions like: "Is it true? Is it true that what I've been taught at home is wrong or is it correct?" And it helps them ask these questions, which enhances learning. They also ask themselves why there's a difference. So this expands their minds in their quest for knowledge."

This stakeholder shows the importance of collaborative learning, learners' freedom to express themselves, and connecting to what learners already know. In turn, the stakeholder explains how this leads to deeper levels of learning: learners asking critical questions and challenging their worldviews.

Follow up 3 learner: How do children use the information they receive during the instructional process?

The ways teachers mentioned learners use information differed in terms of the complexity of information processing by learners. This ranged from reproducing information by imitating or repeating information all the way to challenging the teacher and applying information to real life, both inside and outside of the classroom. The discussion of the answers will be analysed moving from the simplest levels of information use to the most complex levels of information processing.

Of all the teacher answers, the simplest form of information use was one teacher who mentioned learners imitating what they had learned:

"I demonstrate and the learners follow through imitation even outside class."

The second most frequently mentioned use of information was learners practicing or responding to assessments:

"After memorising, they just keep it in order to get the correct answers."

"Whatever you are giving them, when you give them like it's time for class exercise and you notice that they are responding to your objectives, that means they are using the skills they've gotten from what they've been taught."

Both of these forms show rather surface level information use, limited to the classroom context, and most often in response to teacher-initiated activities. The remaining forms of information use were learner-initiated and not limited to the classroom context.

The next form of information use is the third most frequently quoted, which is learners sharing information with others. This can happen both in and outside of the classroom.

"Sometimes it's through stories, they relay the information to their friends. By telling stories like, "This is what we learnt, this is what our teacher has taught us", and such things."

"Sometimes they'll even share it with their families at home."

The next two forms of information use show more complex thinking. They were mentioned less frequently than the previous two forms of information use, but still regularly. One of them is learners relating whatever they learn to their own realities or to existing knowledge:

"Usually when you even just introduce a concept, they also give concrete examples. "You know teacher, it also happened…", "You know when you do that it gives that…". They also bring out their own ideas as long as you've introduced the part, or you've introduced the concept that you want them to learn."

"There are others who can relate from their experiences or from their homes depending on what you've taught them, if it is a familiar information, they'll be able to relate."

The other form is about learners being able to correct or challenge the teacher about what was learned:

"…even correcting the teacher when he/she says a wrong thing about what was taught, which shows that they utilize the information."

"So the fast learners, there is no problem. They can even challenge you as a teacher. You give them something to do, they'll research fast fast, even find related information. They'll go back to you and challenge you by asking those questions… We were talking about oxygen that it is in the air that we breathe in as human beings. And at the same time, oxygen is released by plants. Then afterwards a learner said, "Since you are saying living things use oxygen to breathe like animals and human beings. What about the fish in the water? What happens in the water? Because those animals quite alright have a nose how do they breathe underwater?" Now as a teacher you start thinking nizamuyanka bwanji uyu mwana afunisisa kuziba what happens underwater. So now you start explaining to say even in water there's oxygen and mainly those animals that live in water have gills which they use to breathe when they are underwater."

The last form of information use was the most often quoted. It talks about learners applying their knowledge in real life, changing their behaviour, and creating things as a result of what they have learned. This shows a complex level of information use, which requires learners to not only understand the information they have received, but to recognize new situations and problems in which that information could be used. The excerpts below provide an overview of the different applications of information mentioned by teachers:

"Even when you teach them mathematics, they are able to use the concepts in real life situations, they go to the market. They are able to buy, they are able to change."

"For example, after craft lessons, they are able to make things when they go back to the community. For example, the projects that they learn at school, they will be able to make their own project when they go back home."

"Maybe you have taught them on the environment, on cleanliness. Then during class activities while they are still in class after teaching them, you'll find that one of them has thrown a paper, then the other one will say: "You are not supposed to throw a paper in class. You need to go and throw it outside." Meaning they have responded positively to that lesson you have taught."

This question elicited more elaborate responses from teachers, and interestingly teachers most often referred to deeper forms of learning than to surface level learning. This is different from answers to the previous question. This might have to do with the words used in the two questions: "making sense of information" takes places earlier in the learning process than "using information". It is positive that the majority of teachers focus on learner-initiated use of information, both in and outside of the classroom. The majority of answers focused on a variety of information uses linked to not only the understanding of that information, but to actively use the information when encountering different contexts, problems, and situations.

There was no noteworthy difference between stakeholder responses and teacher responses. Stakeholders also mentioned both surface level uses of information as well as uses that showed a deeper level of understanding by learners. For the surface level, they mentioned learners taking notes as well as responding to questions asked by teachers, which is similar to the teacher responses that

referred to learners using information for assessments and to practice things in class. More complex uses of information were the sharing of information with others, relating information to existing knowledge and real life, and the application of information in daily life.

4 | Conclusion and Discussion

4.1 Conclusion Classroom Observations

For the general implementation of Catch-Up promoted practices, treatment schools scored consistently higher than control schools for all 15 questions included in this section. Five of these differences were found to be significantly different. All these statistically significant differences were found for questions measuring the use of playful activities in class, including songs, dances, games, and stories. These results show a clear indication that in regular classrooms, there is more frequent use of Catch-Up promoted practices in treatment schools than in control schools. These differences are particularly large for playful activities related to the Cheerful component of the 7Cs framework.

For the second section measuring the use of activities reflecting the components of the 7Cs framework, 4 out of 7 components (Concrete, Cheerful, Captivating, and Collaborative) showed consistently higher scores for treatment schools than control schools. For the Concrete component, this difference between treatment and control schools was statistically significant, while for the Cheerful component the difference approached significance. The other two components did not reach statistical significance.

For the Creative component, treatment schools also scored consistently higher than control schools for both the overall scale score and the individual questions, except for the last question on the scale. For the overall scale score, the difference between treatment and control school was not significant at endline but was found to be significant at baseline. Further inspection of the means showed that the size of the difference between treatment and control schools decreased between baseline and endline. It is not clear what has caused this decrease. It could reflect teachers having a better understanding of what the Creative component entailed at endline, and hence they more accurately responded to the questions included.

Results were inconclusive for the Challenging and Connected components. For the Challenging component, the overall scale score was higher for treatment schools than control schools, although this difference did not reach statistical significance. However, for the 4 questions making up the scale, the differences between treatment and control schools were in the expected direction for only 2 questions. For the Connected component, differences between treatment and control schools were negligible for the overall scale score, as well as for 3 out of the 6 questions composing the scale. For the remaining 3 questions, there were observable differences between treatment and control schools, however, only 2 of them were in the expected direction.

Overall, the results of the Classroom Observations show a clear pattern of teachers at treatment schools making use of Catch-Up promoted practices in regular classes more frequently than teachers at control schools. This indicates that there is a spillover of the methods taught in Catch-Up beyond just Catch-Up classes: teachers apply these methods in their regular classes as well. This is especially true for the Concrete, Cheerful, Collaborative and Captivating components of the 7Cs framework. For other components, results remain inconclusive. These findings will be integrated with the results of the qualitative interviews discussed below, as well as with the results of Deliverable 2.

4.2 Conclusion Qualitative Interviews

When teachers describe their role in the classroom setting, imparting knowledge is quoted most frequently. However, teachers also find active involvement of the learners important. Many teachers

refer to themselves as facilitators: someone who provides learners with instructions and guidance to be able to participate in the classroom and engage with the material they are being taught. While teacher answers are mostly focused on learning in general, stakeholders more often also refer to a more holistic development of the child (learning beyond the curriculum) and assuring that no learner is left behind. When it comes to classroom management, supervising and disciplining learners and providing structured and conducive environments are the most frequently mentioned roles for teachers. Occasionally, teachers recognize how their practices can influence holistic development of the learner and assuring that no-one is left behind in the classroom.

For teachers, the most cited effective teaching method was learner centred teaching, after which playful learning came second. Making things practical, using learning and teaching materials, and group work were also mentioned frequently. In addition, stakeholders also made mention of using the learners' language of play. Teachers often explained what the beneficial results were of the suggested teaching methods. They most often cited the promotion of understanding and making sure shy and slow learners participate. Besides these benefits, active engagement of learners, enjoyment, and peer-to-peer learning was also frequently mentioned. This gives us an idea of how teachers understand "effective teaching". Stakeholders' answers were very similar to what was reported by teachers in terms of beneficial outcomes. It is positive that teachers and stakeholders are able to mention a diverse range of activities that can engage learners, different from the traditional copying from the board and memorisation teaching methods. In addition, the benefits they recognize as a result from these methods are aligned with what the Catch-Up program seeks to achieve.

When teachers were asked about which activities can engage their learners in the classroom, teachers mentioned activities that reflected the characteristics of the 7Cs framework and that showed a high level of learner involvement. Some teachers explicitly reported how these activities are more effective than traditional methods, and one third of the interviewees mentioned playful activities. Stakeholder responses were similar to teacher responses, but more high level, referring to characteristics that engaging activities should have instead of naming specific activities. After a general question about which activities could engage their learners, interviewees received multiple probes for sharing activities that included specific characteristics of the 7Cs framework. Answers to these probes show that teachers were able to mention a broad range of diverse activities. After specific probing, stakeholders mentioned the importance of using appropriate teaching and learning materials, allowing learners to engage in free play, and promoting learners to share ideas freely, even when they are not congruent with the content of the curriculum. Their answers show a clear understanding of the 7Cs framework and learning through play.

Most teachers see a passive role for the learner in the classroom setting. When it is a more active role, it is often still in relation to what the teacher instructed learners to do. However, one third of the interviews did mention more active roles for students, and occasionally it is recognized that this is beneficial for the learning process. Stakeholders had similar responses to teachers but had more of a focus on rule setting being a collaborative process between teachers and learners. In addition, they see an opportunity for learners to be actively engaged in classroom management as group leaders. In terms of the learners' role in classroom management, the focus on passive roles for the learner remains. Importantly, some of the interview excerpts show that passive roles for learners do sometimes entail a certain level of agency for learners, more than might be expected at first. This should be further unraveled in future research. It is also important to contextualize this finding into the Catch-Up program, the Zambian culture, and the young ages of the learners involved in the program.

Teachers were also asked how learners make sense of the information they receive in class. More than half of the teachers focused on strategies that were initiated by them, while the rest of them mentioned strategies initiated by learners themselves. This shows that teachers differ in the amount of agency they attribute to learners in the learning process. Many of the strategies that were mentioned were the same as what teachers mentioned for effective teaching methods. Most teacher answers are

not elaborate enough to get a better understanding of how teachers view the learning process, however, when they are, they are most often focused on learners not forgetting what they are taught. This reflects surface level learning: the memorization of knowledge. Stakeholder answers more frequently elaborated on how the suggested strategies relate to the learning process, and these elaborations show that they see learning as more than surface level.

Teachers were also asked how learners use information. For this question, teachers provided more elaborate responses than for the previous question. Most of the answers showed deeper levels of learning: learners applying knowledge in real life, changing behaviour, and creating things. This shows that teachers do not view the learning process as only related to surface level learning. They might see the first phase of learning, when learners make sense of the information, as a process in which learners focus on retaining information, after which they can start actively using and applying it. Future research should ask teachers directly how they view the learning process to get more clarity on this. It is positive that the majority of teachers focus on learner-initiated use of information, both in and outside of the classroom. Most answers focused on a variety of information uses linked to not only the understanding of that information, but the active use and application of information when encountering different contexts, problems, and situations.

In summary, the results of the qualitative interviews show that teachers have a good and elaborate understanding of teaching methodologies and learning strategies that can engage learners and promote learning. Teachers have a wide array of methodologies at their disposal and can explain what the beneficial outcomes are of these methodologies. Their methodologies reflect characteristics of learning through play as explained in the 7Cs framework, but it is important to note that some of the components are represented more often in their responses than others. In general, they are focused on activities, methodologies, and strategies that learners enjoy, allow them to interact with the things they are learning about, and activities that take place in groups. This strongly relates to the Concrete, Cheerful, and Collaborative components of the 7Cs framework. Teachers also often refer to activities that keep learners engaged, even when they would normally be distracted and even those learners that are normally not participating, which relates to the Captivating component. Occasionally, teachers mention activities that relate to the Creative and Connected components, but much less frequent than happens for the aforementioned components. The Challenging component does not appear to be reflected in teachers' answers. This an interesting finding, that will be integrated with findings in other sections and across deliverables. Stakeholder answers more often reflected activities that relate to the Creative, Connected, and Challenging components, which shows their thorough understanding of the 7Cs framework and what it entails.

4.3 Discussion Classroom Observations and Qualitative Interviews

This discussion will look at how the results of the classroom observations and the qualitative interviews relate to each other. All in all, the results of these two sections align. Classroom observations found consistent differences between treatment and control schools for the Concrete, Cheerful, Captivating and Collaborative components. When describing effective teaching and learning methodologies and activities in qualitative interviews, teachers most often mentioned methodologies and activities that reflected characteristics of the Concrete, Cheerful, and Collaborative components. The Captivating component was also often reflected in their answers, albeit a bit less frequent than the other three components. In addition, classroom observations found inconclusive results for the Challenging and Connected component. In qualitative interviews, effective teaching and learning methodologies and activities mentioned by teachers would occasionally reflect characteristics of the Connected component, but the Challenging component seemed absent in the answers they provided. Interestingly, stakeholder answers would regularly reflect characteristics of these components. This shows a discrepancy between stakeholders' understanding of the 7Cs and how it should be implemented and how teachers reported they implement it. When it

comes to the Creative component, the classroom observations did find higher scores for treatment schools compared to control schools on the overall scale score and all items, except for one. However, the difference between treatment and control schools appeared to have reduced between baseline and endline implementation. In qualitative interviews, the Creative component was reflected in the answers of stakeholders but did not occur that often in teacher answers.

Thus, the pattern of results found in classroom observations is reflected in the responses of teachers in qualitative interviews. These concurrent findings suggest that participation in the Catch-Up program had impacts on the frequency teachers use Concrete, Cheerful, Captivating and Collaborative aspects of learning through play in their classroom practices. Impacts on the use of Challenging, Connected, and Creative aspects of learning through play seem limited, even though stakeholders do seem aware of these components and how they could be implemented in the classroom. The reduction in the difference between treatment and control schools on the use of Creative aspects in their practices from baseline to endline cannot be further explained by information gathered in the qualitative interviews and should thus be investigated further.

The results of the current study raise questions about why impact was found on some components of learning through play, but not on others. Different reasons could explain this. One could be that certain aspects of learning through play require less effort to be implemented than others. Cheerful activities like singing, dancing, games, storytelling, and roleplay, as well as Collaborative activities in which learners work together in groups are relatively easy to implement, and do not require a lot of preparation from the side of the teacher. Activities that provoke joy in students and allow them to collaborate are often inherently captivating and stimulate engagement from all learners throughout the activity, thereby relating to the Captivating component. However, this cannot fully explain the differential impacts between the components. Concrete activities for example, which allow learners to directly interact with the things they are learning about, will often require teachers to prepare resources or practical activities. Nevertheless, this was one of the components for which impacts were found in both classroom observations and qualitative interviews.

Another explanation could be that the implementation of some of the components elicit immediate positive results, whilst for others benefits take longer to manifest, and might initially even be negative. Consequently, teachers might need more convincing before implementing those components. For example, teachers reported that the use of Cheerful, Captivating, Collaborative and Concrete components cause feelings of joy among learners, as well as promote their participation, thereby resulting in immediately observable positive impacts. On the other hand, the use of Challenging activities might be difficult to get accustomed to for learners who are used to receiving more guidance from teachers. The lack of guidance and the resulting mistakes and errors learners will make along the way could be particularly discouraging for slower and insecure learners, leading to frustration. Similarly, Creative activities could be unusual for learners used to more controlled classroom environments in which the sharing of their opinions is normally discouraged. This might make learners feel uncomfortable to suddenly start expressing themselves freely, and consequently lead to an initial reduction in learner participation. One teacher in the interviews described that this is one of the benefits of group work. They explained that since some teachers do not create the right environment for learners to share their ideas, this consequently results in learners not feeling at ease sharing things with any teacher. Observing these initial phases of frustration and unease among learners might lead teachers to abandon methods related to the Challenging and Creative components before positive results can take place.

Alternatively, this could also mean that it takes more time for teachers to learn how to effectively implement these activities and create the right classroom atmosphere, since they are so different from traditional methods. Catch-Up refresher training is particularly important in this case, since it gives teachers the opportunity to refresh their knowledge, improve their skills, and find solutions for issues they face after the original training has taken place. However, in another study on the Catch-Up program in Zambia, teachers reported several issues concerning these trainings, including some

teachers not being invited, or refresher trainings being organized too far in the future. These problems could hamper teachers' ability to further master more complicated components of the Catch-Up program, which could translate into less change in their pedagogical practices. Besides that, it is important to reiterate that more than half of the teachers included in the endline were different from teachers selected at those schools for the baseline. Some of these teachers might have not yet participated in Catch-Up training, or only recently, when the endline data collection took place. Hence, for these teachers changes in their pedagogical practices likely had not fully materialized yet.

Lastly, the Creative and Captivating components also require a higher level of agency of learners than the other components do, and qualitative interviews do indicate that the amount of learner agency in classroom activities, as well as making sense of information they receive is relatively limited. Potentially, these components do not align that well with what the Catch-Up program promotes for the young learners included in the program and/or with teachers' ideas on how much agency should be assigned to learners. Future research should investigate whether the current findings can be replicated and should explore what could explain differential impacts on the 7Cs components.

5 | General discussion

This section will focus on the integration of results of Deliverables II and III. It will solely focus on a comparison of results across different deliverables. Findings that were specific to just one of the two deliverables will not be further elaborated on, as this has already been done in the conclusion and discussion sections of the respective deliverables.

First of all, teachers reporting that Cheerful and Collaborative activities happened to a higher extent in their classrooms than teachers at control schools in teacher surveys was congruent with enumerators reporting a higher frequency of the use of these components in teachers' classroom practices for treatment schools compared to control schools. In qualitative interviews, teachers reported teaching methods and activities that also reflected Cheerful and Collaborative components frequently. The replication of these impacts across three different instruments suggests that the Catch-Up program indeed impacted the use of these components in teachers' general classroom practices. Classroom observations and qualitative interviews also found impacts on the Captivating and Concrete components, however, only a small number of questions mapped on these components in the teacher survey (1 question for the Captivating component, and only 2 for the Concrete component), thus this made it difficult to observe consistent patterns for these components in the teacher survey.

A recurring theme in both deliverables is the use of instructional materials. Both teacher surveys and classroom observations included a question about the use of instructional materials adapted to learners' learning needs. In both teacher surveys and classroom observations, the difference in reported use between treatment and control schools was negligible and not significant, however, control schools had higher scores in both instances. These findings relate to two other questions in the teacher survey. One asked teachers if they felt well-equipped for the teaching job, for which teachers at treatment schools scored lower than teachers at control schools. Another question was asked only to teachers at treatment schools, which asked them whether they felt they have the necessary resources to teach the Catch-Up program. Scores were close to the midpoint of the scale, which shows a neutral attitude towards this statement. These responses indicate that resources are likely not sufficient for teachers. This is further corroborated in some of the teacher interviews, which describe that there is a lack of resources to successfully implement the Catch-Up program and that more should be provided. This is also echoed in another study conducted on the Catch-Up program in Zambia, and this study cites the lack of resources as a key barrier to teachers' (quality) implementation of Catch-Up (Busara, 2023). In interviews, both teachers and stakeholders do regularly mention the use of teaching and learning materials as an effective and engaging teaching methodology. Interviewees explain that the use of these materials helps learners make sense of the information they receive and recognize other benefits of instructional materials as well. Stakeholders are among the ones who would be in the position to provide more resources to teachers and schools, so this recognition is positive, but does beg the question why these resources are reported to not be provided sufficiently. It would be of interest to investigate this in future research. Is it that teachers do not use adequate resources because they are not available? Or is it the case that teachers do use resources, but use materials that are not appropriate for the local context or the age of the learners? Since instructional materials are considered an important barrier for Catch-Up implementation and

assure that teaching includes components of learning through play (like Concrete and Connected components), it would be worthwhile to understand how the use of adequate materials can be stimulated.

The patterns of impact found in the implementation fidelity tool, classroom observations, and qualitative interviews was absent in large parts of the teacher survey. Sections measuring teachers' knowledge and practices of the Catch-Up methodology were largely inconclusive and inconsistent. This stark contrast between the teacher survey and other tools is expected to be largely due to methodological issues of the teacher survey, which were exacerbated by the absence of a pilot exercise. Questions included in sections measuring teachers' knowledge and practices regularly showed ambiguous alignment with Catch-Up, and some sentences were lengthy and worded unclearly. Instead of focusing the survey tool at a specific set of practices promoted by Catch-Up, the tool aimed to capture a wide range of Catch-Up promoted activities, to explore which of the Catch-Up practices showed patterns of impact and which did not. However, casting such a wide net had the downfall that most Catch-Up practices were measured by a singular question. In combination with the alignment and wording issues mentioned above, variance on singular questions was likely largely influenced by random noise, instead of treatment effects. Future research should investigate whether addressing these methodological issues could lead to results more in line with the other results.

It should also be noted that while clear patterns of results appeared across deliverables, most of these patterns did not reach statistical significance. Providing robust tests of statistical significance was not necessarily the aim of the current study, which had a more exploratory nature. However, it is still important to reflect on the absence of statistical significance, and what this means for the results of this study. As can be seen in the Annex 1 and 2, the study was underpowered for most of the statistical tests performed, especially for the teacher survey. This relates to the effect sizes of the performed statistical analyses being relatively small, meaning that while differences between treatment and control schools are observable, they are not large enough to reach statistical significance with the sample size of the current study. This could have something to do with the way Catch-Up training for teachers takes place. As mentioned in Deliverable II, Catch-Up training takes place over a course of 4 days, after which periodical refresher trainings are organized. However, research on the Catch-Up program in Zambia shows that the organization of these refresher trainings faces some issues (Busara, 2023). Some teachers report never being invited to these trainings, while others state they should be organized earlier on in the process (Busara, 2023). In general, teachers echoed the need for more and frequent refresher trainings (Busara, 2023). Literature on quality teacher professional development states that when development activities are hosted over a short period of time, follow ups are necessary for these activities to promote intellectual and pedagogical change (Desimone, 2009; Merchie, Tuytens, Devos, & Vanderlinde, 2016). The absence of desired changes in teachers' knowledge and practices could thus be related to issues concerning the necessary refresher trainings.

In addition, the current study focused on spillover effects of the Catch-Up program onto general classroom practices of teachers. Teacher surveys and qualitative interviews asked questions to teachers about their teaching ideas and practices without referring to Catch-Up, and classroom observations were conducted in regular classrooms. While the implementation fidelity observations were conducted specifically in Catch-Up classes, the tool was relatively short and only included as a quick check to see if implementation took place. Questions in the tool measured the adherence to general Catch-Up guidelines. The tool was not developed to extensively measure the use of Catch-Up promoted classroom practices, like the classroom observations tool was. Having a more elaborate measure of teachers' classroom practices in Catch-Up classes would have enabled us to determine whether some of the inconclusive results and small effect sizes in teacher surveys and classroom

observations could be explained by teachers' not using (some) Catch-Up promoted. The leap to measure spillover effects of Catch-Up in general classrooms might have been too large of a jump. In this discussion, it is important to remember that at 10 schools implementation fidelity observations could not be conducted, since Catch-Up teachers were not present at school and/or Catch-Up classes were not organized at the day of data collection. These schools might have been schools with a lower Catch-Up implementation fidelity, which could have influenced the results of relatively high implementation fidelity that were currently found.

Besides that, teachers participating in the study were not asked about their level of involvement in the Catch-Up program. As mentioned before, Catch-Up is implemented at the class level, not the teacher level. This means that if a teacher changes classes within a school, the Catch-Up intervention remains with the class he/she originally taught. Consequently, the new teacher assigned to this class has to take over the Catch-Up program. In the classroom observations, 46 schools out of 77 schools included had different teachers sampled for baseline and endline. It is possible that for these 46 schools, some of the teachers who participated in the endline data collection had only very recently or not yet participated in a Catch-Up training. For this group of teachers, impacts of the Catch-Up program might not have manifested yet given the recency of the training. Anecdotal evidence from both the research team and the VVOB program staff also indicates that some teachers report that there were delays in implementation of the program, or that they were not invited for refresher trainings. In the current study, this could have resulted in an underestimation of the effects. Future research should more elaborately measure teachers' actual involvement in the program, including questions about participation in the Catch-Up training, when this training took place, and how long they have already been participating in the Catch-Up program.

All in all, results of the current study do show interesting patterns of differences that suggest impact of the Catch-Up program on teachers teaching and classroom practices. These results elicit questions for future research. Firstly, can results on classroom observations, implementation fidelity, and qualitative interviews be replicated on teacher self-report surveys when an improved teacher survey tool is used? Secondly, it would be of interest to further investigate whether the differential impacts on components of the 7Cs framework could be replicated, as well as determine what causes these differential impacts. Thirdly, research should investigate how certain aspects of teachers' motivation at treatment schools could be retained (feelings of contributing to learning and learners' engagement and joy) and promoted (feeling well-equipped and appropriately compensated). Fourthly, research should aim to get a deeper understanding of some of the topics touched upon in the qualitative interviews, including how teachers view the learning process, as well as how they see the role of the learner in a classroom.

6 | Annexes Deliverables II and III

6.1 Annex 1: Power for ANOVA analyses on Teacher Surveys

Power was computed using the software G*Power version 3.1.9.6 (Faul, Erdfelder, Buchner, & Lang, 2009). "F tests" were selected for *Test family*, "ANOVA: Fixed effects, omnibus, one-way" for *Statistical test*, and "Post hoc" for *Type of power analysis*. Effect size *f* was computed by using the internal G*Power calculation tool, which uses the between and within variances for this calculation. *a err prob* was kept at 0.05 and the *Number of groups* at 2. The table below shows the power for the ANOVAs that were performed in the Teacher Survey analyses. ANOVAs always included treatment status as the independent variable. The study was moderately powered for ANOVAs performed on variables A10, C2, and C7. For the rest of the ANOVAs performed, the study was underpowered.

Dependent variable	Section	Total sample size	Effect size *f*	Power
A1	A	84	0.106	0.16
A2	A	84	0.105	0.16
A3	A	84	0.171	0.34
A4	A	84	0.013	0.05
A5	A	84	0.031	0.06
A6	A	84	0.025	0.06
A7	A	84	0.053	0.08
A8	A	84	0.006	0.05
A9	A	84	0.071	0.10

A10	A	84	0.232	0.56
A11	A	84	0.067	0.09
B1	B	84	0.165	0.32
B2	B	84	0.028	0.06
B3	B	84	0.117	0.18
B4	B	84	0.144	0.26
B5	B	84	0.070	0.10
B6	B	84	0.217	0.50
B7	B	84	0.145	0.26
B8	B	84	0.056	0.08
B9	B	84	0.171	0.34
B10	B	84	0.181	0.38
B11	B	84	0.114	0.18
B12	B	84	0.018	0.05
B13	B	84	0.176	0.36
B14	B	84	0.135	0.23
C1	C	84	0.182	0.38

C2	C	84	0.150	0.27
C3	C	84	0.257	0.64
C4	C	84	0.105	0.16
C5	C	84	0.097	0.14
C6	C	84	0.027	0.06
C7	C	84	0.222	0.52

6.2 Annex 2: Power for ANCOVA analyses on Classroom Observations

Power was computed using the software G*Power version 3.1.9.6 (Faul, Erdfelder, Buchner, & Lang, 2009). "F tests" were selected for *Test family*, "ANCOVA: Fixed effects, main effects and interactions" for *Statistical test*, and "Post hoc" for *Type of power analysis*. Effect size *f* was computed by using the internal G*Power calculation tool, in which the r^2 of the ANCOVA model was input for the partial η^2. *a err prob* was kept at 0.05, *Numerator df* was put at 1, *Number of groups* at 2, and *Number of covariates* at 1. The table below shows the power for the 8 different ANCOVAs that were performed in the classroom observations analyses. ANCOVAs always included treatment status as the independent variable, and the baseline score of the dependent variable as the covariate. The study was underpowered for the ANCOVAs performed for section G and the Cheerful component, moderately powered for the ANCOVAs performed for the Collaborative and Concrete components, and sufficiently powered for the ANCOVA performed for the Creative component.

Dependent variable	**Total sample size**	r^2	**Power**
General section	74	.0645	.61
Cheerful component	77	.0469	.48
Captivating component	77	.0347	.38
Challenging component	77	.0134	.17

Connected component	76	.0029	.08
Collaborative component	77	.0907	.78
Concrete component	77	.0816	.73
Creative component	77	.1432	.94

6.3 Annex 3: Complete ANCOVA tables

ANCOVA analysis for the general section.

Number of observations	74	***R-squared***	0.0645
Root MSE	.716112	***Adj R-squared***	0.0382

Source	Partial SS	Df	MS	*F*	*p*-value
Model	2.5109247	2	1.2554624	2.45	0.094
Treatment	2.5109083	1	2.5109083	4.90	0.030
Baseline score	.25824801	1	.25824801	0.50	0.480
Residual	36.409992	71	.51281679		
Total	38.920917	73	.53316324		

ANCOVA analysis for the Cheerful component.

Number of observations	77	***R-squared***	0.0469
Root MSE	.939498	***Adj R-squared***	0.0212

Source	Partial SS	Df	MS	*F*	*p*-value
Model	3.2173106	2	1.6086553	1.82	0.169
Treatment	3.2121225	1	3.2121225	3.64	0.060
Baseline score	.2437737	1	.2437737	0.28	0.601
Residual	65.316601	74	.88265678		
Total	68.533912	76	.901762		

ANCOVA analysis for the Captivating component.

Number of observations	77	***R-squared***	0.0347
Root MSE	.919521	***Adj R-squared***	0.0086

Source	Partial SS	Df	MS	*F*	*p*-value
Model	2.2461989	2	1.1230995	1.33	0.271
Treatment	2.1425782	1	2.1425782	2.53	0.116
Baseline score	.01116582	1	.01116582	0.01	0.909
Residual	62.568344	74	.84551816		
Total	64.814543	76	.85282293		

ANCOVA analysis for the Challenging component.

Number of observations	77	***R-squared***	0.0134
Root MSE	1.14178	***Adj R-squared***	-0.0133

Source	Partial SS	Df	MS	*F*	*p*-value
Model	1.3111017	2	.65555086	0.50	0.607
Treatment	.42346743	1	.42346743	0.32	0.570
Baseline score	1.1849265	1	1.1849265	0.91	0.344
Residual	96.471366	74	1.3036671		
Total	97.782468	76	1.2866114		

ANCOVA analysis for the Connected component.

Number of observations	76	***R-squared***	0.0029
Root MSE	1.05977	***Adj R-squared***	-0.0244

Source	Partial SS	Df	MS	*F*	*p*-value
Model	.23873329	2	.11936664	0.11	0.899
Treatment	.12373029	1	.12373029	0.11	0.741
Baseline score	.05471125	1	.05471125	0.05	0.826
Residual	81.987874	73	1.1231216		
Total	82.226608	75	1.0963548		

ANCOVA analysis for the Collaborative component.

Number of observations	77	***R-squared***	0.0907
Root MSE	1.03944	***Adj R-squared***	0.0662

Source	Partial SS	Df	MS	*F*	*p*-value

Model	7.9780391	2	3.9890195	3.69	0.030
Treatment	1.8609448	1	1.8609448	1.72	0.193
Baseline score	4.9822955	1	4.9822955	4.61	0.035
Residual	79.952156	74	1.0804345		
Total	87.930195	76	1.1569762		

ANCOVA analysis for the Concrete component.

Number of observations	77	***R-squared***	0.0816
Root MSE	.916522	***Adj R-squared***	0.0568

Source	Partial SS	Df	MS	*F*	*p*-value
Model	5.5257282	2	2.7628641	3.29	0.043
Treatment	3.408287	1	3.408287	4.06	0.048
Baseline score	1.3244735	1	1.3244735	1.58	0.213
Residual	62.16096	74	.84001297		
Total	67.686688	76	.89061432		

ANCOVA analysis for the Creative component.

Number of observations	77	***R-squared***	0.1432
Root MSE	1.02069	***Adj R-squared***	0.1200

Source	Partial SS	Df	MS	*F*	*p*-value
Model	12.88378	2	6.4418902	6.18	0.003
Treatment	.00817887	1	.00817887	0.01	0.930
Baseline score	11.923656	1	11.923656	11.45	0.001
Residual	77.093133	74	1.0417991		
Total	89.976913	76	1.1839068		

Instructions to enumerator: This section of the questionnaire is to be administered to Teachers in Lusaka province (treatment group). Use it as an observation tool (for CATCH – UP classes) to evaluate the existence of the activities/items indicated below by indicating the appropriate section in the rating scale provided.

6.4 Annex 4: Teacher Instruments

1. Implementation Fidelity Tool

Questionnaire/Tool for Teachers in Lusaka province

- *Based on the Catch-Up training that you have received, rate the teacher's implementation of the core Catch Up methods and Learning Through Play characteristics.*
- *Select one response for each question.*

No.	Question	Response options
1.	Learners are grouped by learning level rather than by Grade (based on assessment scores) [Ask the teacher this Question – how were the groups were formed?]	☐ No ☐ Yes
2.	The teacher ensures that activities capture the 7 Cs (*Concrete, Captivating, Connected, Challenging, Collaborative, Creative, Cheerful*)	☐ Not at all ☐ A little bit ☐ Somewhat ☐ A lot ☐ Very much
3.		☐ Not at all ☐ A little bit ☐ Somewhat

	Teacher began the class with a whole class activity/activity.	☐ A lot ☐ Very much
4.	Teacher grouped learners in small groups for an activity.	☐ Not at all ☐ A little bit ☐ Somewhat ☐ A lot ☐ Very much
5.	The teacher allows the group leader (learner leading the group) to lead some classroom activities	☐ Not at all ☐ A little bit ☐ Somewhat ☐ A lot ☐ Very much ☐ Yes ☐ No
6.	Teacher gave learners individual activities.	☐ Not at all ☐ A little bit ☐ Somewhat ☐ A lot ☐ Very much
7.	Teacher receives mentorship/support from school mentors [*ask teacher whether this done*]	☐ No activity ☐ 1 Activity ☐ 2 Activities ☐ 3 Activities ☐ More than 3 activities
8.	During class, the following number of activities were observed	☐ Before regular class ☐ During regular class ☐ After regular class
9.	What time does the CATCH-UP class take place?	☐ < 1 hour ☐ Approximately 1 hour ☐ > than 1 hour

10.	What is the duration of the CATCH-UP class?	

2. Teacher Survey Tool

Consent

Instructions to enumerator: This questionnaire is to be administered to teachers in both Cath-UP and non-Catch – Up schools.Please take note of the following:

Introduce yourself: name and where you are from. Explain to the respondent the purpose of the study. Explain that CAPOLSAis conducting a study (endline survey) on Socioemotional Learning in the Catch-Up program. Inform the respondent that thereis no immediate benefit from participating in this study. The respondents should be informed that their responses will be treated confidential and will be used for research purposes only. Inform the respondent that the interview will take anywherebetween 15 to 30 minutes. Seek their permission to proceed [use consent form / section provided below].

---

1. Do you agree to participate in this study?
    - No
    - Yes

Demographic Characteristics

Enter today's date
    - yyyy-mm-dd

2. Assessor
    - Brian Zulu
    - Lemington Mweemba
    - Nyambe Ndunda
    - Pedzisai Chatora
    - Freckson Chatyoka
    - Viliza Silwamba
    - Sokoni Mutale
    - Annie Hangoma
    - Kalunga Mooya
    - Mbaita Sitali
    - Chipo Hapunda
    - John Chisanga
    - Francisca Chisompola
    - Sandra Kaseke
    - Mweengwe Shafwuluma
    - Doreen Hamooya
    - Chisanga Kalamba
    - Andrew Kamangu
    - Vanessa Musonda
    - Maureen Mungule
    - Chanda Mukolwe
    - Mutinta Niza Shanjina

- Margaret Simwaamba
- Fatima Chisenga
- Linda Jere
- Kapembwa Kangali
- Tiwonge Nadi
- Kalima Kalima
- Haatembo Mooya
- Brighton Sitwala Sikota

A3.1. Province

Lusaka Province

- Lusaka
- Central

A3.2.1 Districts

Lusaka districts

- Luangwa
- Lusaka
- Chilanga

A3.2a. Schools

Luangwa Schools

- Chilukusha primary school
- Kaluluzi primary school

A3.2b Schools

Lusaka schools

- Adonai Tildelis
- Chawama primary
- Vision of God primary
- Destiny community school
- Matshbaldon
- Tomrho community school
- Bantu Bonse
- Muliswa primary
- Harvest community school
- Twashuka primary
- Prince Takamado primary
- Mahopo primary
- Mthunzi Christian school
- Chaisa primary
- Bereen primary school
- Chilukusha primary
- Kaluluzi primary school
- Blue star primary
- Winning ways community
- Rhema mission
- Mutinta academy
- Emedel community
- Umodzi community
- New ng'ombe

- Peniel
- Living faith
- Crossover
- Harry Mwanga Nkumbula
- Daina Kaimba
- Rolita Christian Community school
- Glory of Christ
- Kalingalinga open
- Vera Chiluba
- Ngwelele primary
- Chitanda primary
- Burma road
- St Lawrence primary
- Elishela Community school
- St Francis of Assisi
- Chingwele
- Chaisa Community School
- Mwima Community School
- Worldwide Community School
- Faimo Community School
- Favour Community School
- Sinai Community School
- Kumbaya Community School
- Ngombe Presbyterian Community School
- Sister's Monijoe Community School
- Chinso Community School

A3.2c Schools

Chilanga schools

- Missions of Hope Christian
- Messiah
- St Stephen
- Manaca
- Joseph Conteh
- Guardian angels Community
- Sosco Game & Fisheries
- Chinyanja
- Lusaka West basic
- Mt Makulu
- Linda Open

A3.2.2 Districts

Central Province districts

- Chibombo
- Shibuyunji

A3.2d Schools

Chibombo schools

- Kasensa Community
- Lyansa Community
- Atuyandane Community
- Mukalashi Primary School

- Shampande
- Chibombo Primary School
- Chinwangoba Community
- Chonkuma Community
- Itumbwe Community
- Kalili Mwandwe Community School
- Nsambe Community
- Mukunkwa Community School
- Lubundi Community School
- Mwanobunda Community
- Makoka Primary
- Mululu Community
- Kafululu Primary
- Shimukuni Primary School
- Kalonga Community
- Kayoba Community
- Mukachembe Community School
- Sililo iri Centre
- Chiko Community School
- Katole Community
- St. Theresa Community
- Kapila Primary School
- Katuba Primary School
- Chimana Community School
- Kabemba Community
- Kayonje Community
- Shikomba Community
- Chititi Primary School
- Naluvanda Community
- Sunshine Community
- Dudzai Community
- Mutakwa Primary School
- Chinsanshi Community
- Nkutika Community
- Nthampwila Community
- Musopelo Primary School
- Kamukwesukeni Community
- Kalutyanka Community
- Mayota Primary School
- Mboshya Primary School
- Muchenje Primary School
- Chitendela Community School
- Kasensa Community School
- Maranatha Community
- Mwashinyambu Community
- Chilikwela Community
- Chitanda Lumamba Community
- Mulunda Community
- Shifwankula Primary
- Chipeso Primary

- Namununga Primary
- Kabangalala Primary
- Mauzu Community
- BMC School
- Kanwanduba Community
- Shabwalala Community

A3.2e Schools

Shibuyunji schools

- The Kings Mission Primary
- Shibuyunji Primary
- Makombwe Primary School
- Chisumbu Community
- Buwelu School

SECTION A: Teaching process

THE TEACHING PROCESS: In this section I am going to ask you some questions about what you think about how leaners aretaught. Please indicate how you much you disagree or agree with each of the following statements.

A1. A good teacher demonstrates the correct way to solve a problem.

- Strongly Disagree
- Disagree
- Agree
- Strong Agree

A2. It is better when the teacher - not the learner – decides what activities are done in the classroom.

- Strongly Disagree
- Disagree
- Agree
- Strong Agree

A3. My role as a teacher is to facilitate the learner's own inquiry.

- Strongly Disagree
- Disagree
- Agree
- Strong Agree

A4. Teachers know a lot more than learners therefore they should not let the learners develop answers that may beincorrect when they can just explain the answers directly.

- Strongly Disagree
- Disagree
- Agree
- Strong Agree

A5. Learners learn best by finding solutions to problems on their own.

- Strongly Disagree
- Disagree
- Agree
- Strong Agree

A6. Teaching should be built around problems with clear, correct answers, and around ideas that most learners cangrasp quickly.

- Strongly Disagree
- Disagree
- Agree
- Strong Agree

A7. How much learners learn depends in how much background knowledge they have - that is why teaching factors isso necessary.

- Strongly Disagree
- Disagree
- Agree
- Strong Agree

A8. Learners should be allowed to think of solutions to practical problems themselves before the teacher shows themhow they are solved.

- Strongly Disagree

- Disagree
- Agree
- Strong Agree

A9. When referring to a 'good performance' I mean a performance that lies above the previous achievement of the learner.

- Strongly Disagree
- Disagree
- Agree
- Strong Agree

A10. A quiet classroom is generally needed for effective learning.

- Strongly Disagree
- Disagree
- Agree
- Strong Agree

A11. Thinking and reasoning processes are more important that specific curriculum content.

- Strongly Disagree
- Disagree
- Agree
- Strong Agree

SECTION B: The Learning experiences.

SECTION B: The Learning experiences; In this section I am going to ask you some questions about what you think about how leaners learn. Please answer by showing the extent to which the following happen in your classroom. Answer by indicating 'To a great extent; Somewhat; Very little; Not at all'.

B1. The class includes challenges and puzzles.

- To a great extent
- Somewhat
- Very little
- Not at all

B2. The class includes games, songs, or dances.

- To a great extent
- Somewhat
- Very little
- Not at all

B3. Learners have choices between different activities done in class.

- To a great extent
- Somewhat
- Very little
- Not at all

B4. The class includes hands-on learning experiences.

- To a great extent
- Somewhat
- Very little
- Not at all

B5. Learners receive the opportunity to try again if failed.

- To a great extent
- Somewhat
- Very little
- Not at all

B6. The teachers give away cues and hints to the students as to help them answering a question.

- To a great extent
- Somewhat
- Very little
- Not at all

B7. Learners receive relevant activities in class to the subject taught.

- To a great extent
- Somewhat
- Very little
- Not at all

B8. I support learners in class using instructional materials adapted to their learning needs.

- To a great extent
- Somewhat
- Very little
- Not at all

B9. Learners had to work in small groups.

- To a great extent
- Somewhat
- Very little
- Not at all

B10. The class includes positive interactions between the children.

- To a great extent
- Somewhat
- Very little
- Not at all

B11. The class activities include songs and games.

- To a great extent
- Somewhat
- Very little
- Not at all

B12. The class activities are interactive.

- To a great extent
- Somewhat
- Very little
- Not at all

B13. Learners are allowed to share their feelings and emotions freely in class.

- To a great extent
- Somewhat
- Very little
- Not at all

B14. I deliberately make an effort to build relationships with my learners.

- To a great extent
- Somewhat
- Very little
- Not at all

SECTION C: Teacher Motivation

SECTION C: Teacher Motivation In this section, I am going to ask you questions about yourself as a teacher in your school.Please indicate how much you disagree or agree with the following statements as they relate to you.

C1. All in all, I am satisfied with my job.

- Strongly disagree
- Disagree
- Agree
- Strongly Agree

C2. My job helps me to meet my basic needs e.g., water, food etc.

- Strongly disagree
- Disagree
- Agree
- Strongly Agree

C3. My job pays me well.

- Strongly disagree
- Disagree
- Agree
- Strongly Agree

C4. I consider myself well equipped for teaching.

- Strongly disagree
- Disagree
- Agree
- Strongly Agree

C5. I feel that I am making a significant education difference in the lives of my learners.

- Strongly disagree
- Disagree
- Agree
- Strongly Agree

C6. I find satisfaction in seeing my learners acquire knowledge in a fun way.

- Strongly disagree
- Disagree
- Agree
- Strongly Agree

C7. Learners seem to enjoy the way classes are handled through play.

- Strongly disagree
- Disagree
- Agree
- Strongly Agree

SECTION D: Environment and self

SECTION D: Environment and self; In this section, I will ask you about your thoughts and feelings and your environmentaround your participation on Catch – Up. Respond by indicating much you agree/disagree with the statement below?

D1. I do not enjoy teaching catch-up classes.

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

D2. Teaching Catch-Up class is not beneficial to my learners.

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

D3. Other teachers who were also trained on the catch-up program are also enjoying teaching on the catch-upprogram.

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

D4. My school authorities do not approve of my participation on the catch-up program.

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

D5. I feel that I do not have the necessary resources to teach on the catch-up program.

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

D6. In the face of challenges, I have difficulty to improvise so that I can deliver the catch-up class to my learners.

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

D7. I intend to continue teaching catch-up classes for at least the next 1 year.

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

D8. I have not yet started teaching catch-up classes.

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

We have reached the end of this questionnaire. Thank you for your time and participation!
You have reached the end of the questionnaire.
Record your current location.
Get GPS point after finishing the observation.

- latitude (x.y °)
- longitude (x.y °)
- altitude (m)
- accuracy (m)

3. Classroom Observations Baseline Tool
Consent
No assessor is to administer the assessment without teacher consent. If the teacher withdraws consent at any point, document the reason for withdrawal with the QCO. If consent is not given, thank the teacher and end the assessment. Assessors are to read the verbal consent script clearly to the teacher to provide them with detailed information about the nature of the research study, its purpose and the data protection protocols surrounding their participation.

• Hello, my name is _____.
• My colleagues and I are working with VVOB and CAPOLSA to conduct assessments of the outcomes of the CATCH UP programmes. This includes an assessment of learners' social and emotional skills, as well as an observation of classrooms. We are also gathering additional information about the Grade three, four and five classroom environment that may influence children learning.
• This school was randomly selected for participation in this research. You are being invited to participate because your experience as a CATCH UP classroom teacher can help inform the Ministry of General Education. Your participation is very important, but you do not have to participate if you do not wish to.
• If you agree to participate, I will ask you some questions regarding your normal activities at school. My questions for you will take approximately 5-10 minutes. In addition, I will observe your classroom for one day/session, taking note of a normal day of classes. I will not ask the learners questions but rather will observe the environment quietly with your consent.
• Your name will NOT be recorded on this form, nor mentioned anywhere in the survey data. The combined results of all the classroom observations conducted in many schools will be shared with VVOB, MoGE and other education stakeholders. They will use the results to identify areas where additional support may be needed to improve early Grade learning.
• You will not personally benefit from participating in this interview or observation. However, your responses will be used to help support improvements in primary education in Zambia.
• If you have any questions regarding this research, please ask me or contact the Principal Investigator.
• Once again, you do not have to participate if you do not wish to. Once we begin, if you would rather not answer a question, that's all right. Do you have any questions? Are you willing to participate?

Will you help us? ☐ YES ☐ NO

Allow the teacher to ask any questions or seek clarification after the statement is written. Once the script is read, and if the teacher provides verbal consent to participate, have him or her complete the *Statement by the researcher/person taking consent* and proceed to *Section I. School & Observation Information.*

Teacher names are not collected to protect the teacher's individual identify. Teachers will be asked to

provide verbal consent. No teacher signature is required, the assessor is only to mark whether they responded with a [YES] or [NO] and then KoboCollect will automatically record the date and time.

1.1. Spill over CU to regular classes

Following questions are directed at the interviewer who observes XX regular class.

G1. The teacher sings songs with the children.
- Never
- Sometimes
- Regularly
- Often
- Very often

G2. The teacher dances with the children.
- Never
- Sometimes
- Regularly
- Often
- Very often

G3. The teacher plays word games with the children.
- Never
- Sometimes
- Regularly
- Often
- Very often

G4. The teacher plays number games with the children.
- Never
- Sometimes
- Regularly
- Often
- Very often

G5. The teacher organises small group activities among the children.
- Never
- Sometimes
- Regularly
- Often
- Very often

G6. The teacher encourages the children to tell a story to the other children in small groups or class.
- Never
- Sometimes
- Regularly
- Often

- Very often

G7. The teacher encourages the children to take the lead in small groups or class.
- Never
- Sometimes
- Regularly
- Often
- Very often

G8. The teacher encourages the child to find the answer to the question.
- Never
- Sometimes
- Regularly
- Often
- Very often

G9. The teacher encourages the child to ask another child the answer on a question when he/she does not know the answer.
- Never
- Sometimes
- Regularly
- Often
- Very often

G10. The teacher creates a welcoming atmosphere for all children.
- Never
- Sometimes
- Regularly
- Often
- Very often

G11.The teacher talks with the children in the mother tongue.
- Never
- Sometimes
- Regularly
- Often
- Very often

G12. The teacher talks with the children on things that are not related to a course or exam.
- Never
- Sometimes
- Regularly
- Often
- Very often

G13. The teacher encourages the children to express his/her feelings.
- Never
- Sometimes
- Regularly
- Often
- Very often

G14. The children laugh in class.
- Never
- Sometimes
- Regularly
- Often
- Very often

1.2. Classroom observation: LtP characteristics of a regular class

• *Carefully read the picture below. The picture contains a checklist of characteristics of learning through play which can be applied in regular classes.*

• *Having had the opportunity to attend a regular class, what is your overall judgement, as an interviewer, on the activities that took place in this class?*

H1. The activities in this class were cheerful?
- Not at all
- A little bit
- Somewhat
- A lot
- Very much

H1.1 The class included challenges and puzzles.
- Not at all
- A little bit
- Somewhat
- A lot
- Very much

H1.2 The class included games, songs or dances.
- Not at all
- A little bit
- Somewhat
- A lot
- Very much

H1.3 The class contained activities that asked from the children to use all senses.
- Not at all
- A little bit
- Somewhat
- A lot
- Very much

Write down in a few words your key observations regarding the characteristic joyful.

H2. The activities in this class were captivating?
- Not at all
- A little bit

- Somewhat
- A lot
- Very much

H2.1 Students had choices between different activities done in class.

- Not at all
- A little bit
- Somewhat
- A lot
- Very much

H2.2 The class included hands-on learning experiences.

- Not at all
- A little bit
- Somewhat
- A lot
- Very much

H2.3 The teacher asked questions to the students.

- Not at all
- A little bit
- Somewhat
- A lot
- Very much

Write down in a few words your key observations regarding the characteristic captivating.

H3. The activities in this class were challenging.

- Not at all
- A little bit
- Somewhat
- A lot
- Very much

H3.1 Learners received the opportunity to try again if failed.

- Not at all
- A little bit
- Somewhat
- A lot
- Very much

H3.2 The teachers gave away cues and hints to the students as to help them answering a question.

- Not at all
- A little bit
- Somewhat
- A lot
- Very much

H3.3 The teacher helped students by giving them appropriate ways of solving a question.

- Not at all
- A little bit
- Somewhat
- A lot
- Very much

Write down in a few words your key observations regarding the characteristic challenging.

H4. The activities in this class were connected?

- Not at all
- A little bit
- Somewhat
- A lot
- Very much

H4.1 Students received relevant activities in class to the subject taught.

- Not at all
- A little bit

- Somewhat
- A lot
- Very much

H4.2 The teacher used templates, structures, or examples in class.

- Not at all
- A little bit
- Somewhat
- A lot
- Very much

H4.3 The teacher supported students in class using instruction materials adapted to the learning needs of the students.

- Not at all
- A little bit
- Somewhat
- A lot
- Very much

Write down in a few words your key observations regarding the characteristic connected.

H5. The activities in this class were collaborative?

- Not at all
- A little bit
- Somewhat
- A lot
- Very much

H5.1 Students had to work in small groups.

- Not at all
- A little bit
- Somewhat
- A lot
- Very much

H5.2 The class included positive interactions between children.

- Not at all
- A little bit
- Somewhat
- A lot
- Very much

H5.3 The class included positive interactions between the teacher and the children.

- Not at all
- A little bit
- Somewhat
- A lot
- Very much

H5.4 Students had to present their work to the other children in class.

- Not at all
- A little bit
- Somewhat
- A lot
- Very much

Write down in a few words your key observations regarding the characteristic collaborative.

H6. The activities in this class were concrete?

- Not at all
- A little bit
- Somewhat
- A lot
- Very much

H6.1 The class activities included songs and games.

- Not at all

- A little bit
- Somewhat
- A lot
- Very much

H6.2 The class included story telling activities.

- Not at all
- A little bit
- Somewhat
- A lot
- Very much

H6.3 The classroom activities were interactive.

- Not at all
- A little bit
- Somewhat
- A lot
- Very much

Write down in a few words your key observations regarding the characteristic concrete.

H7. The activities in this class were creative?

- Not at all
- A little bit
- Somewhat
- A lot
- Very much

H7.1 Learners were allowed to share their feelings and emotions freely.

- Not at all
- A little bit
- Somewhat
- A lot
- Very much

H7.2 The classroom environment was open and comfortable.

- Not at all
- A little bit
- Somewhat
- A lot
- Very much

H7.3 The teacher made an effort to build relationships with the learners.

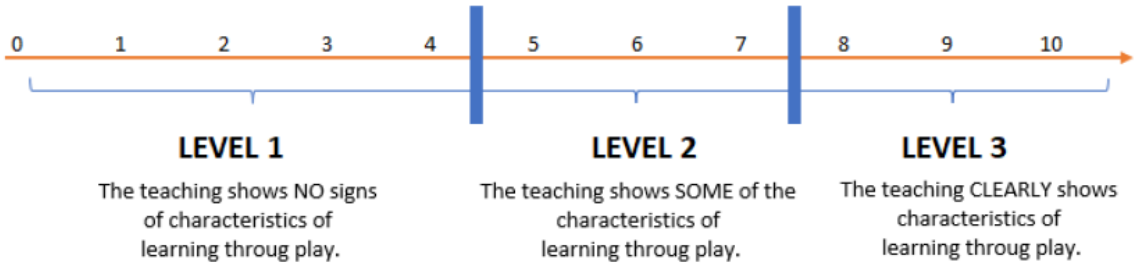
- Not at all
- A little bit
- Somewhat
- A lot
- Very much

H7.4 The class conversations were mostly positive.

- Not at all
- A little bit
- Somewhat
- A lot
- Very much

Write down in a few words your key observations regarding the characteristic creative.

• *Based on the analysis and interpretation you just did, please write down your own conclusion:*



because: ……..

4. Classroom Observations Endline Tool
Consent
No assessor is to administer the assessment without teacher consent. If the teacher withdraws consent at any point, document the reason for withdrawal with the QCO. If consent is not given, thank the teacher and end the assessment. Assessors are to read the verbal consent script clearly to the teacher to provide them with detailed information about the nature of the research study, its purpose and the data protection protocols surrounding their participation.

• Hello, my name is _____.
• My colleagues and I are working with VVOB and CAPOLSA to conduct assessments of the outcomes of the CATCH UP programmes. This includes an assessment of learners' social and emotional skills, as

well as an observation of classrooms. We are also gathering additional information about the Grade three, four and five classroom environment that may influence children learning.
• This school was randomly selected for participation in this research. You are being invited to participate because your experience as a CATCH UP classroom teacher can help inform the Ministry of General Education. Your participation is very important, but you do not have to participate if you do not wish to.
• If you agree to participate, I will ask you some questions regarding your normal activities at school. My questions for you will take approximately 5-10 minutes. In addition, I will observe your classroom for one day/session, taking note of a normal day of classes. I will not ask the learners questions but rather will observe the environment quietly with your consent.
• Your name will NOT be recorded on this form, nor mentioned anywhere in the survey data. The combined results of all the classroom observations conducted in many schools will be shared with VVOB, MoGE and other education stakeholders. They will use the results to identify areas where additional support may be needed to improve early Grade learning.
• You will not personally benefit from participating in this interview or observation. However, your responses will be used to help support improvements in primary education in Zambia.
• If you have any questions regarding this research, please ask me or contact the Principal Investigator.
• Once again, you do not have to participate if you do not wish to. Once we begin, if you would rather not answer a question, that's all right. Do you have any questions? Are you willing to participate?

Will you help us? ☐ YES ☐ NO

Allow the teacher to ask any questions or seek clarification after the statement is written. Once the script is read, and if the teacher provides verbal consent to participate, have him or her complete the *Statement by the researcher/person taking consent* and proceed to *Section I. School & Observation Information.*

Teacher names are not collected to protect the teacher's individual identify. Teachers will be asked to provide verbal consent. No teacher signature is required, the assessor is only to mark whether they responded with a [YES] or [NO] and then KoboCollect will automatically record the date and time.

1.1. Spill over CU to regular classes

G1. The teacher has a lesson plan.
- No
- Yes

G2. The teacher sings songs with the children.
- No
- Yes

G3. The teacher dances with the children.
- Never
- Sometimes
- Regularly

- Often
- Very often

G4. The teacher plays word games with the children.
- Never
- Sometimes
- Regularly
- Often
- Very often

G5. The teacher plays number games with the children.
- Never
- Sometimes
- Regularly
- Often
- Very often

G6. The teacher organises small group activities among the children.
- Never
- Sometimes
- Regularly
- Often
- Very often

G7. The teacher encourages the children to tell a story to the other children in small groups or class.
- Never
- Sometimes
- Regularly
- Often
- Very often

G8. The teacher encourages the children to take the lead in small groups or class.
- Never
- Sometimes
- Regularly
- Often
- Very often

G9. The teacher encourages the child to find the answer to the question.
- Never
- Sometimes
- Regularly
- Often
- Very often

G10. The teacher encourages the child to ask another child the answer on a question when he/she does not know the answer.
- Never
- Sometimes
- Regularly
- Often
- Very often

G11. The teacher creates a welcoming atmosphere for all the children.
- Never
- Sometimes
- Regularly
- Often
- Very often

G12.The teacher talks with the children in the mother tongue.
- Never
- Sometimes
- Regularly
- Often

- Very often

G13. The teacher talks with the children on things that are not related to a course or exam.

- Never
- Sometimes
- Regularly
- Often
- Very often

G14. The teacher encourages the children to express his/her feelings.

- Never
- Sometimes
- Regularly
- Often
- Very often

G15. The children laugh in class.

- Never
- Sometimes
- Regularly
- Often
- Very often

1.2. Classroom observation: LtP characteristics of a regular class

• *Carefully read the picture below. The picture contains a checklist of characteristics of learning through play which can be applied in regular classes.*

• *Having had the opportunity to attend a regular class, what is your overall judgement, as an interviewer, on the activities that took place in this class?*

H1. The activities in this class were cheerful?

H1.1 The class included challenges and puzzles.

- Not at all

- A little bit
- Somewhat
- A lot
- Very much

H1.2 The class included games, songs or dances.

- Not at all
- A little bit
- Somewhat
- A lot
- Very much

H1.3 The class contained activities that asked from the children to use all senses.

- Not at all
- A little bit
- Somewhat
- A lot
- Very much

Write down in a few words your key observations regarding the characteristic Cheerful.

H2. The activities in this class were captivating?

H2.1 Learners engage and participate in classroom activities.

- Not at all
- A little bit
- Somewhat
- A lot
- Very much

H2.2 Learners show interest in classroom activities.

- Not at all
- A little bit
- Somewhat
- A lot
- Very much

H2.3 Learners had choices between different activities done in class.

- Not at all
- A little bit
- Somewhat
- A lot
- Very much

H2.4 The class included hands-on learning experiences.

- Not at all
- A little bit
- Somewhat
- A lot
- Very much

H2.5 The teacher asked questions to the students.

- Not at all
- A little bit
- Somewhat
- A lot
- Very much

Write down in a few words your key observations regarding the characteristic captivating.

H3. The activities in this class were challenging.

H3.1 Learners help each other to solve challenges and complete tasks given by teachers.

- Not at all
- A little bit
- Somewhat
- A lot
- Very much

H3.2 Learners received the opportunity to try again if failed.
- Not at all
- A little bit
- Somewhat
- A lot
- Very much

H3.3 The teachers gave away cues and hints to the students as to help them answering a question.
- Not at all
- A little bit
- Somewhat
- A lot
- Very much

H3.4 The teacher helped students by giving them appropriate ways of solving a question.
- Not at all
- A little bit
- Somewhat
- A lot
- Very much

Write down in a few words your key observations regarding the characteristic challenging.

H4. The activities in this class were connected?

H4.1 The teacher refers to the local context of the learners in his/her lessons.
- Not at all
- A little bit
- Somewhat
- A lot
- Very much

H4.2 The teacher uses examples from real life or from previous lessons.
- Not at all
- A little bit
- Somewhat
- A lot
- Very much

H4.3 Learners received relevant activities in class to the subject taught.
- Not at all
- A little bit
- Somewhat
- A lot
- Very much

H4.4 The teacher used templates, structures, or examples in class from previous lessons.
- Not at all
- A little bit
- Somewhat
- A lot
- Very much

H4.5 The teacher supported learners in class using instruction materials adapted to the learning needs of the learners.
- Not at all
- A little bit
- Somewhat
- A lot
- Very much

H4.6 The teacher uses familiar language in his/her interaction with the learners.
- Not at all
- A little bit
- Somewhat
- A lot

- Very much

Write down in a few words your key observations regarding the characteristic connected.

H5. The activities in this class were collaborative?

H5.1 Students had to work in small groups.

- Not at all
- A little bit
- Somewhat
- A lot
- Very much

H5.2 The class included positive interactions between children.

- Not at all
- A little bit
- Somewhat
- A lot
- Very much

H5.3 The class included positive interactions between the teacher and the children.

- Not at all
- A little bit
- Somewhat
- A lot
- Very much

H5.4 Students had to present their work to the other children in class.

- Not at all
- A little bit
- Somewhat
- A lot
- Very much

Write down in a few words your key observations regarding the characteristic collaborative.

H6. The activities in this class were concrete?

H6.1 The teacher uses concrete materials such as stones, sticks, baskets, etc.

- Not at all
- A little bit
- Somewhat
- A lot
- Very much

H6.2 The class activities included songs and games.

- Not at all
- A little bit
- Somewhat
- A lot
- Very much

H6.3 The class included story telling activities.

- Not at all
- A little bit
- Somewhat
- A lot
- Very much

H6.4 The classroom activities were interactive.

- Not at all
- A little bit
- Somewhat
- A lot
- Very much

Write down in a few words your key observations regarding the characteristic concrete.

H7. The activities in this class were creative?

H7.1 Learners were allowed to share their feelings and emotions freely.

- Not at all
- A little bit
- Somewhat
- A lot
- Very much

H7.2 The classroom environment was open and comfortable.

- Not at all
- A little bit
- Somewhat
- A lot
- Very much

H7.3 The teacher made an effort to build relationships with the learners.

- Not at all
- A little bit
- Somewhat
- A lot
- Very much

H7.4 The class conversations were mostly positive.

……

Conclusion and Rating

Below are three levels for the rating (Scale 1-10 & their meaning) of the observation you have made.
Read carefully.

LEVEL ONE (1-4)
The teacher shows NO sign of characteristics of learning trough play.
LEVEL TWO (5-7)
The teacher shows SOME of the characteristics of learning trough play.
LEVEL THREE (8-10)
The teacher CLEARLY shows characteristics of learning through play.

Based on the analysis and interpretation you just did, please select your own conclusion:

- *LEVEL ONE (1-4)*
- *LEVEL TWO (5-7)*
- *LEVEL THREE (8-10)*

What's the reason for your selection?
Briefly explain.

……..

This is the end of the observation. Thank you very much.
Record your current location.
Get GPS point after finishing the observation.

- latitude (x.y °)
- longitude (x.y °)
- altitude (m)
- accuracy (m)

5. Qualitative Interview Tool

Instructions to enumerator: This is an in-depth Teacher and Stakeholder questionnaire. Ensure that you administer the questionnaire in a quiet place. Please take note of the following:

1. Introduce yourself: name and where you are from.

2. Explain to the respondent the purpose of the study. Explain that CAPOLSA is conducting a study on Socioemotional Learning in the Catch-Up program. Inform the respondent that there is no immediate benefits for participating in this study.

4. The respondents should be informed that their responses will be treated confidentially and will be used for research purposes only. The respondents should be encouraged to express their views freely. Participants should be reminded that there are no 'wrong or right answers' and no one will judge them for the responses they will give. Emphasise that this is not an exam or an assessment.

5. Please ensure that all you cover, in as much as possible, the questions in this instrument, unless the respondent refuses to. Probe where necessary. While the structure of the questions is important, try to ensure a smooth flow in the conversation.

7. Inform the respondent that the interview will take anywhere between 15 to 30 minutes.

Seek their permission to proceed [use consent form / section provided below]

_____

1. Would you want to participate in this interview?

    1. No ☐
    2. Yes ☐

2. Name of interviewer: _____ Date: _____ /_____ / 2022

3. Are you working with [relevant only for stakeholder interviews]

- Ministry of Education ☐
- VVOB ☐
- TARL Africa ☐

SECTION A: Interview questions

#	Question
A1	**Teacher** From your experience, what do you think is the role of the teacher in a classroom setting?

	○ Probe ○ Teacher's role in classroom management? ○ What do you think is the most effective way to teach children? ○ What activities can the teacher engage his/her learners in to enhance their participation in class? Probe about activities that 1. are engaging; 2. attract learner's interests; 3. relate to learner's (already existing) interests; 4. appeal to learner's creativity and invoke imagination; 5. allow learner's freedom to express themselves; 6. encourage learners to work with others; and 7. bring joy to the children.
A2	**Learner** From your experience, what do you think is the role of the learner in a classroom setting? ○ Probe ○ What strategies do you think children use (best) to make sense of the information they receive in class e.g., *do they memorise; apply the knowledge* etc? ○ How do children use the information they receive during the instructional process? ○ What is the learner's role in classroom management?
Thank you for your participation	

References

Ajzen, I. (1991). The theory of planned behavior. *Organizational behavior and human decision processes*, *50*(2), 179-211.

Busara Center for Behavioral Economics (2023). Barriers and enablers to quality Catch Up delivery in Zambia. [Unpublished paper].

Das, B. L., & Baruah, M. (2013). Employee retention: A review of literature. *Journal of business and management, 14*(2), 8-16.

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational researcher, 38*(3), 181-199.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149-1160.

Gokce, F. (2010). Assessment of teacher motivation. *School Leadership and Management*, *30*(5), 487-499.

Kaymakamoglu, S., E. (2018). Teachers' Beliefs, Perceived Practice and Actual Classroom Practice in Relation to Traditional (Teacher-Centered) and Constructivist (Learner Centered) Teaching (Note 1). *Journal of Education and Learning*, *7*(1), 29-37.

Loeb, S., & Luczak, L. D. H. (2013). How teaching conditions predict: Teacher turnover in California schools. In *Rendering school resources more effective* (pp. 48-99). Routledge.

Merchie, E., Tuytens, M., Devos, G., & Vanderlinde, R. (2018). Evaluating teachers' professional development initiatives: towards an extended evaluative framework. *Research papers in education, 33*(2), 143-168.

Ministry of Education, Republic of Zambia (2020). *Chapter 10: Learning through play at ECE*. https://zambia.vvob.org/sites/zambia/files/10._learning_through_play_lf.pdf

Nduku, T., Staskowicz, E., & Stern, J. (2022). Playful Learning Across the Years (PLAY) Measurement Toolkit: Full Report.

Organisation for Economic Co-operation and Development (OECD) (2018). *Teaching and Learning International Survey (TALIS) 2018: Teacher Questionnaire, Main Survey Version*. https://web-archive.oecd.org/2020-04-30/499238-TALIS-2018-MS-Teacher-Questionnaire-ENG.pdf

Zosh, J. N., Hopkins, E. J., Jensen, H., Liu, C., Neale, D., Hirsh-Pasek, K., & Whitebread, D. (2017). *Learning through play: a review of the evidence*. White Paper, Billund, Denmark: LEGO Foundation.