Unlocking the Black Box: To what extent are interactive classrooms effective classrooms in Andhra Pradesh and Telangana, India?

1.0 Introduction

Successful completion of secondary education is crucial in breaking the intergenerational transmission of poverty (World Bank, 2009). Yet, while access to secondary school is increasing across India, its potential benefits are only partially realised, in part owing to limitations of school quality. There is extensive evidence of poor learning outcomes among a majority of students and of weak mastery of the skills required by 21st century economies (ASER, 2018; ASER, 2017). Nonetheless, a small minority of students in India demonstrate learning outcomes comparable to those in OECD countries (Das & Zajonc, 2010) while the context is one of highly unequal access to 'quality' education. 'Sorting' of disadvantaged pupils into poor quality schools is a driver of inequality and inequity in educational opportunities, with potentially far-reaching consequences for wider social and economic disparities.

While it is clear that the nature of interactions in the classroom are central to educational quality, there is little large-scale evidence on this. Existing research suggests that broad summary indicators of classroom activity such as 'time on task' or 'opportunity to learn' (see USAID, 2010) are predictive of students' learning outcomes and progress, including in low and middle-income countries. The leap from summary indicators of 'time and effort' to 'quality' measures of classroom interactions is challenging since it requires first a normative framework for quality judgements, and second a team of appropriately skilled trained observers. Until recently, large-scale studies focused on quality of classroom interactions have been rare outside of developed countries.

In this paper, we draw on new data from two states in India to examine classroom interactions and their relationships to student learning and the 'sorting' of students into schools, alongside their equity implications. Data are drawn from the Young Lives longitudinal study which, in India, covers Andhra Pradesh (AP) and Telangana, which are considered 'median states' in terms of both their human development index and net state domestic product (Kapoor, 2017).

The paper addresses the following research questions:

1. To what extent do 'more effective' teachers (estimated in terms of teacher value-added) employ strategies which score highly using the CLASS (Classroom Assessment Scoring System) observation protocol?
2. Who are the teachers achieving higher scores on the CLASS protocol and whom do they teach?
3. Under what conditions can a relationship between learning progress and teacher practices (as identified by CLASS) be observed?

We identify a weakly positive association between 'teacher effectiveness' and 'teacher quality' (measured using CLASS), with a stronger relationship in certain types of schools. We discuss reasons for this, and the implications for policy makers.

2.0 Inequality and inequity in the Indian education system

The expansion of mass education in India has placed strain on the existing system, worsened by the limited resources the government commits to the costs of educational provision (Kingdon, 2017). Existing studies suggest that the forces of competition and 'school choice' have led to quality of schooling being 'allocated' according to discriminatory factors such as ability to pay school fees or access to information about school quality, as well as contextual factors, such as geography (Härmä, 2011; James & Woodhead, 2014). Children are 'sorted' into differentially effective schools and classes according to their backgrounds, to the detriment of girls, those from poorer families, and children with less educated parents (Anonymous, 2018). Analyses in this paper provide further depth to this evidence of 'sorting', indicating that lower achieving children in poorly performing government schools are less likely to be taught by teachers using interactive or supportive pedagogies, with the implication that existing inequalities in learning outcomes will continue to widen.

2.1 Teacher Effectiveness and Value-Added

Readily available indicators of educational quality provide limited proxies for effective classroom processes. Teacher education, qualifications, experience and pay, for example, typically show very weak explanatory power in statistical modelling exercises (see Glewwe 2011; Hanushek and Rivkin, 2012), although some studies in India do identify a modest effect for teacher training, professional knowledge, and union membership (Azam & Kingdon, 2015; Singh & Sarkar, 2015; Kingdon & Muzzamil, 2013; Bhattacharjea et al, 2011). Inconsistency among results nonetheless means there is no straightforward way to identify 'effective' teachers based on background characteristics. Classroom-level 'educational production' remains something of a 'black box', partly because of data limitations but also because 'what matters' depends on the setting and context.

Alternative approaches focus on summary measures of the 'whole' contribution made by teachers to students' learning progress, such as 'teacher value-added'. Value-added approaches broadly equate educational quality with the learning gains made by a group of students during a defined period (Ladd, 2008), either with or without adjusting for student backgrounds. Repeated measures of student learning permit modelling exercises that account for differences in student baseline scores and therefore compare 'like with like' where learning progress is concerned (Perry, 2016). Debate surrounds the precision and stability of value-added estimates (McCaffrey et al, 2004) while nonetheless their use is increasingly common.

Value-added analysis typically identifies substantial variation between teachers, including those within the same schools (see Rivkin et al, 2005; Araujo et al, 2016), with a small number of developing country studies now available (see Aslam et al, 2019). In India, Azam & Kingdon (2015) estimate teacher value-added using panel data from private schools in Uttar Pradesh and identify large variation in teacher performance within and between schools, remarking that "the same student can systematically score significantly different marks given different teacher quality" (ibid:82). Value-added analysis using Young Lives data in India finds similar variation in teacher effectiveness (Anonymous, 2018). Yet while differences at the teacher level can explain a large proportion of variation in student test scores in these studies in South Asia, consistently identifying the characteristics of 'more effective' teachers remains challenging (Muralidharan and Sundararaman, 2011; Aslam and Rawal, 2015; Singh and Sarkar, 2015).

2.2 Understanding teacher effects through classroom observation

Classroom observation has a long history in educational research, most often in small-scale research studies, as well as in school improvement efforts (Goe et al, 2008). Given the difficulties in identifying simple indicators of teacher effectiveness, classroom observation approaches remain important for their explanatory power, not least in developing countries[1] (Bruns et al, 2016). For example, recent studies using CLASS (Classroom Assessment Scoring System) in Ecuador, Chile and the USA find that improvements in teachers' observed classroom practices are linked to increases in students' test scores (Rockoff, 2004; Araujo et al, 2016; Bruns et al, 2018).

In India, several classroom observation studies have taken place in recent years. Two of these were conducted by the World Bank: one at primary level (2006-7), and another at secondary (2015-16). Both used the Stallings method (Stallings, 1980), which records time spent by teachers and students on different activities at 'snapshot' time-points (World Bank, 2014). 'Traditional', teacher-led activities such as lecture and demonstration were found to dominate in these Indian studies (World Bank, 2014; 2016), with students typically spending relatively little time on 'higher order activities' requiring more critical and independent thinking. Young Lives' Stallings-based observation study which took place in 2010 found somewhat similar results (Galab et al, 2014); with small-group work and other cooperative student-led activities observed in less than 1 percent of sampled classes (Singh & Sarkar, 2015). These studies found strong evidence of student engagement in lessons, with little sign of poor discipline or students being 'off-task' (World Bank, 2014; 2016); however this was not understood as unambiguously positive, as "when all students were involved this almost always meant that all students were doing the same activity, regardless of whether they were able to understand what was being taught" (World Bank, 2014: 6). A large-scale longitudinal study conducted by ASER offers a more nuanced picture of this finding, highlighting that 'child

[1] The World Bank's *Teach* observation tool, launched late in 2018, offers an alternative approach to classroom observation aimed specifically at low and middle-income countries (Molina et al, 2018). This tool was not available at the time of the Young Lives study.

friendly' classrooms which truly engage children and support learning require a combination of "classroom environment, teacher attitude, teaching methods, and student activities" (Bhattacharjea et al, 2011: 36), rather than strong classroom management alone.

The use of more sophisticated methods than Stallings is increasing in developing countries (Bruns et al, 2016). This includes studies using CLASS, which has been shown to predict student learning as well as other, non-cognitive, outcomes (Downer et al, 2010; Hamre et al, 2014; Pianta and Hamre, 2009a). While studies show that a teacher's CLASS score is predictive of students' outcomes, this need not be a causal relationship. More able, advantaged or motivated students may be allocated to higher scoring CLASS teachers, or CLASS scores may be correlated with other features of good teaching which are not measured directly by CLASS, such as subject knowledge. Some studies attempt to address these sources of confounding, but this remains a challenge.

3.0 Data and Methodology

In this paper, we draw on large-scale secondary school effectiveness data collected in two Indian states in 2016-17 and on classroom observation data from a linked sub-study (2017-18). The school-level data were collected from classes in Grade 9. Data were collected from four school types: state government, tribal/social welfare, private unaided and private aided. Schools in the effectiveness study are in seven districts, while the observation study included a sub-sample of urban and rural schools in four districts.

We adopt a sequential mixed methods research strategy focused on 'expansion' and deepening of insights (Greene et al, 1989) gained through combining one method (survey data analysis) with another (classroom observation). While data collected from both methods are quantitative, classroom observation 'scores' are based on observer judgements. These judgements are made against an established rubric (CLASS), which essentially provides a means of quantifying a qualitative process.

3.1 Young Lives and Young Lives School Surveys

Young Lives has followed the lives of 12,000 children in Ethiopia, India, Peru and Vietnam since 2001. 20 purposively selected sites are included in each country to represent national diversity. In India these are all located within AP and Telangana. The Young Lives secondary school survey took place in 2016-17 in 205 schools sampled within 20 *mandal*-level sites, and collected data from almost 10,000 Grade 9 students, 519 maths and English teachers and 205 head teachers (Anonymous, 2017). We employ data from this survey alongside data from a linked classroom observation study.

3.2 Young Lives Classroom Observation Study design

The Young Lives classroom observation study was conducted in 2017-18 in a sub-set of schools from the larger 2016-17 school survey. The study aimed to strengthen and deepen understanding of the classroom environment and of teacher-student interactions for teachers for whom 'value-added' estimates were available. It aimed to examine whether classroom interactions differ substantially and meaningfully between classes with lower and higher levels of learning progress. The study made use of the Classroom Assessment Scoring System-Secondary (CLASS-S) observational tool, which records information on the quality of teacher-student interactions in the classroom.

3.2.1 Classroom Observation using Classroom Assessment Scoring System (CLASS-S)

CLASS-S measures three domains of teacher 'quality': the generation of a warm and supportive classroom environment ('emotional support'); the provision of behavioural support ('classroom organisation'); and the promotion of higher order thinking skills ('instructional support') (Landry et al, 2009). Eleven dimensions fall under these three domains. Table 1 summarises this and provides an example of types of focus areas covered by each dimension. In addition, CLASS-S assesses student engagement through a twelfth standalone dimension which measures students' level of academic involvement and motivation in the classroom (Pianta et al, 2012).

The CLASS tool is notable for its consistent theoretical framework and robust accumulation of empirical evidence (Hu et al, 2016). Its conceptual framework is grounded in theories of effective teaching, attachment, self-determination, responsivity and scaffolding (Bornstein et al, 2008; Bronfenbrenner and Morris, 1998), and posits teacher-student interactions as the primary engine through which children learn (Pianta and Hamre, 2009c). The importance CLASS places on measuring teacher-student interactions stems from the fact that these have been shown to support children's learning (Downer et al, 2010; Curby et al, 2009b), and to help children develop social competencies (Hamre et al, 2014; Curby et al, 2009a;), strengthen long-term academic outcomes (Pianta and Hamre, 2009a, 2009b), and reinforce 'life readiness' skills (Duncan et al, 2013; Heckman and Kautz, 2012).

CLASS and its conceptual framework have been successfully used in many countries, cultures and societies (Hu et al, 2016). There is now considerable evidence to support the applicability of CLASS in both developed and developing countries, for example in Chile (Leyva et al, 2015), Germany (von Suchodoletz et al, 2014), Finland (Pakarinen et al, 2010) Portugal (Cadima et al, 2010) and China (Hu et al, 2016).

[insert Table 1]

3.2.2 Sampling and data collection

Young Lives' classroom observation study employed a purposive sampling strategy, selecting more and less 'effective' teachers using value-added estimates within urban and rural schools and different types of school management (see Table 2). Value-added estimates were estimated using repeated measures (including a lagged, or prior attainment, measure) of student learning in maths and English from the 2016-17 Young Lives school survey within a multilevel regression model (Goldstein, 1997).

[insert Table 2]

Owing to the costly nature of this type of study, the sample was limited to approximately 10 percent of teachers from the 2016-17 school survey: 45 maths and English teachers in 23 schools[2] across the four school types. Characteristics of these schools are summarised in Table 3, showing that private unaided schools typically are larger overall but with smaller class sizes and higher test scores. These findings are common with other studies (for example, ASER, 2018), although those which take into account differences in student background typically find more mixed results on test-scores (Muralidharan and Sundararaman, 2015; Singh, 2015)[3].

[insert Table 3]

Within each school, selected teachers were observed teaching Grade 9 students for two periods[4]. Each teacher was assessed against each CLASS dimension by two trained and

---

[2] None of the teachers within this sample taught both maths and English; it is therefore not possible to look for 'within-teacher' variation as they were only observed teaching one subject.
[3] India has seen a huge growth in 'low fee' private schools in recent decades. However even in these schools, students are more typically advantaged than those attending public schools (Härmä, 2011; Singh & Bangay, 2014).
[4] In total teachers were observed for four 15 minute 'cycles'. The score for each dimension was an average of the scores given across the four cycles.

certified observers[5], and given a score ranging from 1 to 7, as well as an average score for each domain. Based on evidence from other studies (for example Pianta et al, 2012) it was expected that teacher-student interactions could be expected to remain somewhat constant over time, although we were not able to test this assumption owing to this study's sequential design (value-added estimates were based on the previous year's cohort of students).

3.3 Validation of CLASS in the Indian context

CLASS-S has not previously been used in India and results are of course in part a reflection of context. Previous studies in India find evidence of strong 'classroom management' alongside apparently high levels of 'on-task' behaviour among students, reflecting a notable deference to the teacher's authority (World Bank, 2014; 2016) linked in part to historic construction of the teacher as a 'guru' figure (NCTE, 2010: 18).

To examine the validity of the CLASS-S tool, we first conducted an informal qualitative exercise with the trained local observers. We then explored the reliability and internal validity of results from the CLASS-S tool using confirmatory factor analysis. This revealed that the factor structure of the data does not readily conform to the hypothesised three-factor model of CLASS-S established by studies conducted elsewhere (Pianta et al, 2009). Our analysis finds that the three CLASS domains are not readily separable and are strongly correlated (Figure 1). This suggests that teachers who score highly in one domain are likely to score highly in the other two, meaning that the data are largely unidimensional. Accordingly, a single-factor model provides a better fit to the data, so in this study, our analyses focus on an overall CLASS score combining teacher performance across the eleven dimensions.

[5] Observers were recruited from a pool of education doctoral students from AP and Telangana. They were trained and certified using CLASS protocols, which involved a three-day intensive training course followed by an online certification test.

[insert Figure 1]

3.4 Data Analysis

In the sections below, we first examine an overview of the data for the three CLASS domains overall and for selected categories of teachers (Table 4)[6]. Analysis then moves to consider a 'combined CLASS score' which is simply the sum of each teacher's standardised scores[7] for the 11 CLASS dimensions, weighting each dimension equally as all have the same mean and standard deviation. Given the apparent unidimensionality of the data (described above) and the lack of any obvious reason for weighting the dimension scores, this approach was considered reasonable.

[insert Table 4]

We first examine the relationship between teacher classroom practice and 'unconditional' teacher value-added. Teacher value-added is estimated using Young Lives secondary school survey data in a simple two-level (multilevel) model; an established method of generating value-added estimates in educational effectiveness research (see Goldstein, 1997; Goldstein et al, 2000). Using graphical representations of these linked data, we explore the extent to which those teachers identified as 'more effective' are rated more highly on the CLASS protocol; and proceed to examine the results considering the purposive nature of the sample.

We proceed to estimate three models aimed at addressing the research questions in inferential, although not necessarily causal, terms, employing a multilevel design which both

[6] Domain-level scores are simply an average of the scores each teacher received for the separate dimensions assessed by CLASS. The dimensions within each domain and examples of behaviours within each are shown in Table 1.

[7] The score for each dimension is centred on 0 (mean), with a standard deviation of 1.

recognises and explores the effects of students' clustering within teachers and schools (Steele, 2008). All multilevel models used in our analysis are two-level random intercept models, with students at the first level, and teachers at the second. Accordingly, teacher effects can be expected to absorb the unaccounted-for effect of higher-level groups, such as the school[8]. The dependent variable for all three models is student maths and English attainment at the end of Grade 9 (see Table 3 for mean test scores overall and by school type).

The first model explores the question of student selection, examining the extent to which students with higher test scores are 'selected' into classes taught by higher 'quality' teachers (i.e. those with higher scores on the CLASS protocol).

A second model builds on this to estimate the relationship between a teacher's CLASS score and their students' learning progress: a value-added model. This is employed to examine whether, for both maths and English, students learn more over the course of one school year if they are taught by a teacher with a higher CLASS score, taking student 'sorting' into account.

The final model examines the conditions under which a relationship between teacher classroom practice and student learning can be observed. It includes interaction terms to allow relationships between teacher CLASS score and student outcomes to differ between school types. Recognising the considerable complexities of the Indian education system, this model allows us to estimate the differential effects of teacher classroom practice on student learning in different settings.

4.0 Classroom Interactions and Effective Teaching: Descriptive Findings

[8] This is in contrast to the use of a three-level model when drawing the teacher sample for the classroom observation study, where it was desirable to estimate class-level and school-level value-added separately to allow identification of high performing teachers in lower performing schools, and *vice versa*.

Sampled teachers in AP and Telangana demonstrated a wide range of classroom practices (Table 4, above), with particularly high variation on Emotional and Instructional Support. Somewhat less variation was observed in the domain of Classroom Organisation for all teachers observed (whether English/maths, male/female, urban/rural, and across different school types), which includes Behaviour Management, Productivity and Negative Climate. This very general finding aligns with other studies in India, which report evidence of relative consistency in classroom management, linked to comparatively high levels of deference towards the teacher (World Bank 2014; 2016). The high level of consistency in teacher scores for Classroom Organisation indicates that sampled teachers employ largely similar behaviour management strategies. More variation is seen for Instructional Support, and among English teachers, female teachers, and those teaching in state government schools.

Scores on all three domains are positively associated with student learning progress (value-added) (Figure 2), suggesting that the most effective teachers are, in general, those using more interactive methods across the three domains. Lower levels of Emotional and Instructional Support are more prevalent among teachers who are relatively 'less effective'. This supports the idea that teaching practices related to these domains may be targeted to improve learning outcomes.

[insert Figure 2]

Figure 2 also confirms the strong correlation between CLASS domains. For this reason, we use a single combined CLASS score for the analyses that follow. Figure 3 presents teacher-level data for this combined score, with bars grouped by effectiveness as measured in terms of unconditional value-added (recalling that most teachers were sampled partly on this basis). CLASS scores are centred on zero, meaning an estimate of zero represents the mean CLASS score. Similar to the fitted lines shown in Figure 2, this figure reveals a weakly positive association between 'teacher effectiveness' and 'teacher quality'. In particular, teachers with lower than average effectiveness are more often evaluated as lower scorers on CLASS, suggesting that for these teachers in particular, changes in teaching practice might have a positive effect on their students' learning. Nonetheless, there is a considerable degree of overlap between more and less effective teachers, suggesting that the relationship between the interactions measured by CLASS and value-added is not a straightforward one.

[insert Figure 3]

Figure 4 presents the same data as above while highlighting reasons for teachers' selection. The pattern of association is somewhat stronger once teachers who were included in the study for 'another reason' are excluded (as shown in the fitted line in Figure 4). With teachers sampled for the sub-study on the basis of their value-added estimate, this association between reason for sampling and CLASS score indicates some consistency in the two datasets, suggesting a potential relationship between teacher practices and the extent of student learning[9]. However, the relationship is not especially strong. A small number of highly effective teachers do not score highly on CLASS (whether in terms of the combined score or individual domains) and equally a small number of less effective teachers score above average on CLASS. Clearly, there are many routes to improving student learning as measured by test scores, perhaps depending on the aims of the school or teacher: these may be exam driven, teaching to the test using 'rote' learning methods; or alternatively may be more holistic and aimed at a 'rounded education' or at critical thinking. The CLASS methodology may not

[9] It is also possible that both relate to a third, omitted variable which could influence them both, such as teacher content knowledge.

capture everything about teacher effectiveness in these Indian states, but this analysis suggests it can be useful in highlighting hypotheses for further investigation, as well as areas of major concern such as those schools where both CLASS score and value-added are low.

[insert Figure 4]

Figure 5 shows the results by school type. Separate fitted lines show the relationship between teacher CLASS score and value-added estimates for teachers in government and private unaided schools (the two largest categories). A positive relationship is observed for teachers in state government schools, driven largely, it appears, by two groups of teachers – one 'low CLASS-score, low-effectiveness' cluster in the bottom left quadrant, plus a smaller number of teachers with (fairly weakly) correlated effectiveness and CLASS scores in the top right quadrant. Conversely, for the albeit small number of teachers in private unaided schools, the pattern is somewhat less clear and appears be a negative relationship so that more effective teachers achieved lower CLASS scores. This is potentially of interest for further investigation, suggesting that these teachers may be doing something other than using 'interactive' teaching methods to support students; for example, such as focusing on rote learning and techniques for passing exams. However, it should be noted that this relates to a very small sample of private unaided school teachers, all of whom were more effective than the 'average' teacher[10].

[insert Figure 5]

[10] This is in part an effect of sampling: the sample frame included a very small number of private unaided teachers identified as 'less effective' in the selected districts, none of whom were available to participate in the observations. It should also be remembered that, in this study, teacher effectiveness is estimated using an unconditional value-added approach, which controls for student prior attainment but not student background; this benefits teachers in private unaided schools more than other school types.

While there appears to be a positive association overall between teacher effectiveness and performance on the CLASS scale, driven largely by teachers working within state government schools, there are multiple ways in which to be an effective teacher, and different ways in which student progress can be achieved. The considerably stronger pattern seen in government schools may indicate that in these settings what a teacher does in the classroom is closely associated with student learning outcomes, perhaps because these learners have fewer other inputs to support their study. In such settings, it is suggested that the role of the teacher may become even more vital in driving learning. Conversely, in the small number of private unaided schools sampled, we see a much less clear pattern suggesting that although all of these teachers can be classed as effective (with above average value-added), their classroom practices are highly varied, with some utilising strategies which achieve a low score in CLASS yet which appear to deliver high levels of student learning. Of most concern, in some state government schools our analysis suggests that children in lower performing classes are subject to a double disadvantage, which sees them taught by less effective teachers who are also using less supportive classroom practices.

4.1 Learning and Classroom Practices: Multilevel Modelling Results

We employ a series of multilevel models to consider these questions further. Models have been estimated separately for students' maths and English learning attainment, with the results presented in the Appendix (Table A1).

4.1.1 Model 1: Cross-Sectional Predictors of Learning Outcomes

Model 1 includes just one measure of student learning (the outcome) so cannot be considered a measure of teacher 'effectiveness'. Rather, it tells us about the factors that predict student attainment for this small sample of students at one point in time, including those relating to their teachers. The results show a positive and significant association between teacher CLASS score and student test scores at the end of Grade 9 in both maths and English. This indicates that higher achieving children are taught by 'higher quality' teachers, conditioning on differences in school management type and student backgrounds. Because indicators for school type are included, the results show that, overall, within each school type, higher scoring pupils are taught by teachers who use more interactive approaches on average (although as discussed below, Model 3 indicates that this is not the case within all school types once prior attainment is taken into account).

The results of Model 1 are highly suggestive of 'sorting' of students into classes and schools, consistent with existing research on India (Aslam et al, 2019). Other possibilities remain, however, such as the suggestion that, when teaching more able students, teachers may adopt more interactive methods. Alternatively, it is possible that it is teachers rather than students who are 'sorted', with 'higher quality' teachers directed (through state deployment or private school recruitment processes) towards those students who are higher achieving.

4.1.2 Model 2: Teacher Value-Added

Model 2 extends Model 1 into an 'effectiveness' model, by including prior attainment measures. Teachers' CLASS scores remain positively and significantly associated with student learning for English, while for maths the association is no longer significant. With the inclusion of prior attainment, this suggests that, in English, a teacher with a higher CLASS score is associated with increased student progress in Grade 9.

Model results indicate that a one-point increase in an English teacher's total CLASS score (which is equal to a teacher being one standard deviation above average for one of the eleven dimensions) is associated with a 1.5 point increase in their students' English test scores at the end of Grade 9, controlling for prior attainment level. (For context, existing analysis of Young Lives secondary school data shows that an average student makes 10 points of progress in one school year in English (Anonymous, 2017)). There is a gap of 29 points between the combined CLASS score of teachers at the 5th and 95th percentile; this indicates the potential for an association between CLASS score and student learning to have an effect on outcomes (albeit only for English teachers and students within this sample).

4.1.3 Model 3: Heterogeneous Relationships? An Interaction Model

As Figure 5 indicated, there is an apparent interaction between the CLASS - value-added association and type of school, with the indication of different relationships between teacher classroom practices and student learning in government and private schools. Accordingly, Model 3 includes a set of interaction variables to more thoroughly examine this heterogeneity.

In private unaided schools, the interaction between CLASS score and school type is significant and negative for both maths and English in comparison to the state government school reference group, indicating that those private schools in which students are learning most are not necessarily those in which teachers are using pedagogies scoring highly on CLASS. The same pattern is also seen for tribal/social welfare school teachers for English. These complex results suggest a need for additional research in this context on the relationships between teaching and learning strategies effective at raising learning attainment and those considered ‘good practice’ internationally.

5.0 Discussion

Findings from this paper provide further support for claims that children in AP and Telangana are 'sorted' into schools according to their background, offering evidence that higher achieving students are selected into classes taught by teachers scoring more highly on CLASS, or perhaps alternately that higher quality teachers are directed towards those students deemed likely to succeed. It is important to remember that CLASS scores and value-added estimates are measuring very different phenomena. Nonetheless, it appears important where possible for students to have access to a teacher who is both effective and who makes use of positive classroom practices, for both overlapping and distinct reasons.

This paper contributes to the discussion on learning outcomes in India through the finding that those students taught by teachers who engage in more positive classroom practices make more progress in English. This indicates a risk of a widening gap in attainment as lower achieving children (at the class-level[11]) are taught by teachers who are less likely to use interactive or supportive pedagogies, and as a result, are placed at risk of falling further behind. While findings from our analyses reveal that a teacher's CLASS score is not always associated with greater student learning, those schools where children are taught by neither an effective teacher nor one scoring highly on CLASS tend to be mostly poorly performing government schools, and it is these which should be the focus of policy attention. This finding has parallels with earlier studies in India (e.g. Bhattacharjea et al, 2011), which reveal an association between 'child friendly' classrooms (based both on teacher attitudes and their practices) and student learning. Children taught by teachers who are both less effective and who make use of less positive classroom interactions to support their teaching are most likely

[11] Our analysis focuses on understanding differences between teachers and classrooms, rather than differences in teacher-student interactions within a class, as teachers' CLASS scores relate to their interactions with all students.

to benefit from reforms, and indeed must be the focus of attention if the learning crisis in these two states (and more broadly in India) is to be meaningfully addressed.

This paper also draws attention to the complexity in identifying what makes some teachers more effective than others in the Indian context, a finding highlighted in other teacher effectiveness literature from South Asia (Aslam et al, 2019). Within private unaided schools this might be a focus on exams, meaning teachers are prioritising ‘teaching to the test’ rather than more interactive classroom practices; it could also imply other factors are supporting learning in these contexts, such as private tuition. This merits further exploration if we are to gain a full understanding of how those teacher classroom practices measured by CLASS can really be said to impact on student learning in AP and Telangana, India.

References

Araujo, C., Carneiro, P., Cruz-Aguayo, Y., and Schady, N. (2016). Teacher Quality and Learning Outcomes in Kindergarten. IDB-WP-665. Washington DC.: Inter-American Development Bank.

ASER (2017) Annual Status of Education Report (Rural) 2016. New Delhi: ASER Centre.

ASER (2018) 'Beyond Basics': Annual Status of Education Report 2017 (Rural). New Delhi: ASER.

Aslam, M. and Rawal, S., 2015. Teachers - an indispensible asset: examining teacher effectiveness in South Asia. In: P. Dixon, S. Humble and C. Counihan, eds. Handbook of International Development and Education. Cheltenham: Edward Elgar Publishing.pp.256–276.

Aslam, M., Malik, R., Rawal, S., Rose, P., Vignoles, A. and Whitaker, L., 2019. Methodological lessons on measuring quality teaching in Southern contexts, with a focus on India and Pakistan. Research in Comparative and International Education, [online] 14(1), pp.77–98. Available at: <https://doi.org/10.1177/1745499919828929>.

Azam, M. and G.G. Kingdon, (2015) 'Assessing Teacher Quality in India'. Journal of Development Economics. 117: 74-83.

Bhattacharjea, S., Wadhwa, W. and Banerji, R., 2011. Inside Primary Schools. A study of teaching and learning in rural India. Education and the State in Modern Peru. New Delhi.Boyden, J. and Z. James (2014) 'Schooling, childhood poverty and international development: choices and challenges in a longitudinal study'. Oxford Review of Education, 40:1, 10-29.

Bruns, B., De Gregorio, S., and Taut, S. (2016) Measures of Effective Teaching in Developing Countries. RISE Working Paper 16/009. Oxford: RISE.

Bruns, B., L. Costa, and N. Cunha (2018) Through the Looking Glass: Can Classroom Observation and Coaching Improve Teacher Performance in Brazil? Economics of Education Review, 64(C): 214-250.

Bornstein, M., Tamis-LeMonda, C., Hahn, C., and Haynes, O. (2008). Maternal responsiveness to young children at three ages: Longitudinal analysis of a multidimensional, modular, and specific parenting construct. Developmental Psychology, 44, 867–874. doi:10. 1037/0012-1649.44.3.867.

Bronfenbrenner, U., and Morris, P. (1998). The ecology of developmental processes. In W. Damon & R. M. Lerner (Eds.), Handbook of child psychology: Vol. 1. Theoretical 1029. New York, NY: Wiley.

Cadima, J., Leal, T., & Burchinal, M. (2010). The quality of teacher–student interactions: Associations with first graders' academic and behavioral outcomes. Journal of School Psychology, 48, 457–482. doi:10.1016/j.jsp.2010.09.001.

Curby, T. W., LoCasale-Crouch, J., Konold, T. R., Pianta, R. C., Howes, C., Burchinal, M.,Barbarin, O. (2009a). The relations of observed pre-K classroom quality profiles to children's achievement and social competence. Early Education & Development, 20(2), 346–372. doi:10.1080/10409280802581284

Curby, T. W., Rimm-Kaufman, S. E., & Ponitz, C. C. (2009b). Teacher–child interactions and children's achievement trajectories across kindergarten and first grade. Journal of Educational Psychology, 101(4), 912–925. doi:10.1037/ a0016647

Das, J. and T. Zajonc (2010) 'India shining and Bharat drowning: Comparing two Indian states to the worldwide distribution in mathematics achievement'. Journal of Development Economics 92: 175–187.

Downer, J. T., Sabol, T., & Hamre, B. (2010). Teacher child interactions in the classroom: toward a theory of within and cross-domain links to children's developmental outcomes. Early Education and Development, 21, 699–723.

Duncan, G. J., Dowset, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, et al. (2007). School readiness and later achievement. Developmental Psychology, 43, 1428–1446.

Galab, S., P.P. Reddy, and V.N. Reddy (2014) Classroom Process Teacher Ability and Student Performance: Evidence from School based Component of Young Lives in Undivided Andhra Pradesh. CESS Working Paper 134. Hyderabad: CESS.

Glewwe, P. W., Hanushek, E. A., Humpage, S. D., & Ravina, R. (2011). School resources and educational outcomes in developing countries: A review of the literature from 1990 to 2010. National Bureau of Economic Research.

Goe, L., C. Bell and O. Little (2008) Approaches to Evaluating Teacher Effectiveness: A Research Synthesis. Washington, DC: National Comprehensive Center for Teacher Quality.

Goldstein, H. (1997) 'Methods in School Effectiveness Research.' *School Effectiveness and School Improvement* 8 (4): 369-395.

Goldstein, H., P. Huiqi, Rath, T. and Hill, N. (2000) *The Use of Value Added Information in Judging School Performance.* London: Institute of Education.

Hamre, B. K., Hatfield, B., Pianta, R. C., & Jamil, F. (2014). Evidence for general and domain-specific elements of teacher–child interactions: Associations with preschool children's development. Child Development, 85(3), 1257–1274. doi:10.1111/cdev.2014.85.issue-3.

Hanushek, E. A. and Rivkin, S G. (2012), "The Distribution of Teacher Quality and Implications for Policy," Annual Review of Economics, 4, 131-157.

Härmä, J. (2011) 'Low cost private schooling in India: Is it pro poor and equitable?' *International Journal of Educational Development.* 31: 350-356.

Heckman, J., and T. Kautz. 2012. "Hard Evidence on Soft Skills." Journal of Labor Economics 19(4):451-64.

Hu, B., Fan, X., Gu, C., and Yang, N. (2016). Applicability of the Classroom Assessment Scoring System in Chinese Preschools Based on Psychometric Evidence. Early Education and Development, Vol. 27, No. 5, pg. 714 - 734. doi: http://dx.doi.org/10.1080/10409289.2016.1113069.

James, Z. and Woodhead, M. (2014) 'Choosing and changing schools in India's private and government sectors: Young Lives evidence from Andhra Pradesh'. *Oxford Review of Education.* 40(1): 73-90.

Kapoor, A. (2017) Social Progress Index: States of India 2006-17. Eleven Years of Progress. Report Findings. Gurgaon: Social Progress India.

Kingdon, G. and M. Muzzamil (2013) 'The School Governance Environment in Uttar Pradesh, India: Implications for Teacher Accountability and Effort'. The Journal of Development Studies, 49(2): 251-269.

Kingdon, G. G. (2017). *The private schooling phenomenon in India: A review*. IZA Discussion Paper No. 10612. Bonn: Institute of Labor Economics (IZA).

Ladd, H.F. (2008) 'Teacher Effects: What do we know?' Paper presented at Teacher Quality Conference, Northwestern University, 1st May 2008.

Landry, S., Anthony, J., Swank, P., and Monseque-Bailey, P. (2009). Effectiveness of comprehensive professional development for teachers of at-risk preschoolers. Journal of Educational Psychology, 101, 448–465. doi:10.1037/a0013842.

Leyva, D., Barata, M., Snow, C., Rolla, A., Treviño, E., Yoshikawa, H., and Weiland, C. (2015). Teacher–Child Interactions in Chile and Their Associations With Prekindergarten Outcomes. Child Development, Vol 86, 3, 781–799.

McCaffrey, D.F., D.M. Koretz, J.R. Lockwood, and L.S. Hamilton (2004) The Promise and Peril of Using Value-Added Modeling to Measure Teacher Effectiveness. RAND Research Brief. Santa Monica, CA: RAND Corporation.

Molina, E., S.F. Fatima, A. Ho, C.M. Hurtado, T. Wilichowski and A. Pushparatnam (2018) Measuring Teaching Practices at Scale. Results from the Development and Validation of the Teach Classroom Observation Tool. World Bank Policy Research Working Paper 8653. Washington D.C.: World Bank.

Muralidharan, K. and Sundararaman, V. (2011). Teacher performance pay: Experimental evidence from India. Journal of Political Economy, 119(1), pp.39–77.

Muralidharan, K. and Sundararaman, V. (2015). The aggregate effect of school choice: Evidence From a Two-Stage Experiment in India. The Quarterly Journal of Economics, 130(3), pp.1011–1066.

NCTE (2010) *Draft Code of Professional Ethics for School Teachers*. New Delhi.

Pakarinen, E., Lerkkanen, M.-K., Poikkeus, A.-M., Kiuru, N., Siekkinen, M., Rasku-Puttonen, H., & Nurmi, J.-E. (2010). A validation of the Classroom Assessment Scoring System in Finnish kindergartens. Early Education & Development, 21, 95–124. doi:10.1080/10409280902858764.

Perry, T. (2016) ‘English Value-Added Measures: Examining the Limitations of School Performance Measurement’. British Educational Research Journal. 42(6): 1056-1080.

Pianta, R. C., & Hamre, B. K. (2009a). Classroom processes and positive youth development: Conceptualizing, measuring, and improving the capacity of interactions between teachers and students. New Directions for Youth Development, 121, 33–46. doi:10.1002/yd.v2009:121.

Pianta, R. C., & Hamre, B. K. (2009b). A lot of students and their teachers need support: Using a common framework to observe teacher practices might help. Educational Researcher, 38(7), 546–548. doi:10.3102/0013189X09348786.

Pianta, R., and Hamre, B. (2009c). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. Educational Researcher, 38, 109–119.

Pianta, R., Hmare, B., and Mitz,S. (2012). Classroom Assessment Scoring System Secondary Manual. Charlottesville: Teachstone.

Rivkin, S.G., E.A. Hanushek and J.F. Kain (2005) ‘Teachers, Schools, and Academic Achievement’. Econometrica. 73(2): 417-458.

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. American799 Economic Review, 94, 247–252.

Singh, A. (2015). Private school effects in urban and rural India: Panel estimates at primary and secondary school ages. Journal of Development Economics, [online] 113, pp.16–32.

Singh, R. and Bangay, C. (2014). Low fee private schooling in India - More questions than answers? Observations from the Young Lives longitudinal research in Andhra Pradesh. International Journal of Educational Development, 39, pp.142–150.

Singh, R. and S. Sarkar (2015) ‘Does teaching quality matter? Students learning outcome related to teaching quality in public and private primary schools in India’. International Journal of Educational Development 41: 153–163.

Stallings, J. (1980). Allocated academic learning time revisited, or beyond time on task. *Educational researcher*, *9*(11), 11-16.

Steele, F. (2008) Module 5: Introduction to Multilevel Modelling Concepts. Centre for Multilevel Modelling, University of Bristol: Bristol.

USAID (2010) Opportunity to learn: A high impact strategy for improving educational outcomes in developing countries. USAID Educational Quality Improvement Program (EQUIP2)

von Suchodoletz, A., Fäsche, A., Gunzenhauser, C., & Hamre, B. K. (2014). A typical morning in preschool: Observations of teacher–child interactions in German preschools. Early Childhood Research Quarterly, 29, 509–519. doi:10.1016/j.ecresq.2014.05.010

World Bank (2009) Secondary Education in India: Universalizing Opportunity. Washington DC, USA: World Bank.

World Bank (2014) How much and what kind of teaching is there in elementary education in India? Evidence from three States. South Asia Human Development Sector Report No.67. New Delhi: World Bank.

World Bank (2016) What is happening inside classrooms in Indian secondary schools? A Time on Task Study in Madhya Pradesh and Tamil Nadu. Washington DC: World Bank.

Anonymous (2015) Details omitted for review process.

Anonymous (2017). Details omitted for review process.

Anonymous (2017) Details omitted for review process.

Anonymous (2018) Details omitted for review process.

Appendix Table A1:

[insert Table 5 here]

Tables to be inserted

Table 1: CLASS-S Domains and Dimensions*

Domain	Dimension	Example Areas of Focus
Emotional Support	Positive climate	Relationships within the classroom
	Teacher sensitivity	Awareness and responsiveness to student needs
	Regard for adolescent perspectives	Relevance to current life
Classroom organisation	Behaviour management	Setting expectations
	Productivity	Classroom routines
	Negative climate	Disrespect
Instructional Support	Instructional learning formats	Active facilitation
	Content understanding	Content knowledge and misconceptions
	Analysis and inquiry	Opportunities for students' own application
	Quality of feedback	Scaffolding
	Instructional dialogue	Strategies to facilitate dialogue
Student engagement		Active student engagement

* Sources for information contained in this table: Pianta et al., 2012; Hamre et al., 2014.

Table 2: Teacher Sample for the Young Lives Classroom Observation Study

	School management type				District				Total
Reason for sampling	**Private Aided**	**Private Unaided**	**State Govt**	**Tribal Social Welfare**	**Srikak ulam**	**Anant apur**	**Karim nagar**	**Mahbub nagar**	
Low VA	1	0	12	3	2	3	2	9	16
Average VA	2	0	4	0	4	0	2	0	6
High VA	0	5	2	0	5	1	1	0	7
Another reason*	2	3	8	3	3	5	3	5	16
Teacher Total	5	8	26	6	14	9	8	14	45
School Total	3	4	13	3	7	5	4	7	23

* Another reason' covers teachers sampled as replacements when the selected teacher was not present on the day of the visit and those included because another teacher in their school was purposively selected.

Table 3: Background characteristics of sample schools

	Details of schools included in the Classroom Observation sub-sample					Total for full Young Lives school survey sample
	Private Aided	**Private Unaided**	**State Govt**	**Tribal Social Welfare**	**Total**	
Average class size	38	31	39	39	38	43
Average school size	357	792	320	473	432	541
Average end of Grade 9 maths score	530	589	471	406	541	531
Average end of year Grade 9 English score	519	566	408	394	510	511

Table 4: Mean CLASS Scores by CLASS Domain

		Emotional Support		Classroom Organisation		Instructional Support	
		Mean score	**SD**	**Mean score**	**SD**	**Mean score**	**SD**
Subject	**Maths**	4.5	0.84	5.6	0.61	4.3	0.88
	English	4.2	0.92	5.4	0.71	3.8	1.11
Teacher gender	**Male**	4.18	0.85	5.36	0.63	3.83	0.84
	Female	4.67	0.99	5.72	0.67	4.39	1.29
School type	**Private Aided**	4.0	0.36	5.4	0.69	3.56	0.58
	Private Unaided	4.57	0.83	5.77	0.56	4.1	1.02
	State Govt	4.31	1.01	5.36	0.73	4.06	1.13
	TSW	4.54	0.64	5.76	0.41	4.11	0.91
Locality	**Rural**	4.38	0.96	5.51	0.62	4.06	1.07
	Urban	4.23	0.80	5.38	0.77	3.96	0.93

Table 5: A series of multilevel models for maths and English teachers

	Maths			English		
	Model 1	**Model 2**	**Model 3**	**Model 1**	**Model 2**	**Model 3**
VARIABLES	**Cross-Sectional model**	**VA model**	**Interaction model**	**Cross-Sectional model**	**VA model**	**Interaction model**
Student prior attainment						
Prior attainment (irt_mat1 / irt_eng1)		-0.173	-0.183		0.326	0.306
		(0.204)	(0.204)		(0.207)	(0.207)
Prior attainment^2 (mat2 / eng2)		0.000867***	0.000892***		0.000249	0.000278
		(0.000210)	(0.000210)		(0.000210)	(0.000210)
School and teacher characteristics						
Teacher combined standardised CLASS score	2.880**	0.517	0.731	1.950***	1.511***	2.080***
	(1.443)	(0.649)	(0.625)	(0.716)	(0.436)	(0.350)
School type (ref cat: State Govt)						
Private Aided	32.01	0.510	20.39	6.029	1.989	0.900
	(35.80)	(16.13)	(15.73)	(38.84)	(27.63)	(24.39)
Private Unaided	85.87**	29.23*	40.52***	130.5***	54.67***	52.60***
	(34.38)	(15.60)	(14.51)	(17.98)	(12.18)	(9.588)
Tribal Social Welfare	-63.50*	-39.69**	-41.92***	-11.17	-4.980	-12.02
	(35.86)	(17.53)	(15.40)	(21.13)	(13.63)	(11.44)
Private Aided X Total CLASS score			5.728**			-
			(2.792)			
Private Unaided X Total CLASS score			-2.190*			-2.888***
			(1.225)			(0.987)
Tribal Social Welfare X Total CLASS score			0.965			-3.831***
			(2.154)			(1.358)
Student background characteristics						
Male	-0.789	-9.296	-7.154	-10.16**	-1.722	0.490

	(8.413)	(6.731)	(6.679)	(4.883)	(4.244)	(4.198)
Age	-8.493*	-3.246	-3.450	-3.716	-0.980	-1.279
	(4.996)	(4.000)	(3.979)	(2.910)	(2.495)	(2.455)
Wealth index	3.257	0.307	-0.0627	-1.054	-1.304	-1.837
	(2.506)	(2.042)	(2.074)	(1.557)	(1.340)	(1.342)
Social category (ref cat: General Caste)						
Scheduled Caste (SC)	-48.67***	-31.23**	-26.02**	-6.849	-6.031	-8.399
	(15.45)	(12.63)	(12.86)	(10.32)	(8.859)	(8.771)
Scheduled Tribe (ST)	-55.34**	-26.79	-25.64	-18.45	-19.14	-11.11
	(22.76)	(18.13)	(18.69)	(15.23)	(12.78)	(13.39)
Other Backward Caste (OBC)	-34.01***	-23.65**	-20.13*	-11.15	-6.870	-7.375
	(12.86)	(10.55)	(10.59)	(7.743)	(6.686)	(6.637)
Mother's education (ref cat: No education)						
Primary school	3.815	0.315	1.838	5.367	4.667	4.023
	(8.926)	(7.279)	(7.376)	(5.214)	(4.492)	(4.477)
Secondary school	29.03**	11.12	11.72	19.18***	6.189	5.188
	(11.65)	(9.502)	(9.648)	(6.860)	(5.976)	(5.937)
Higher education	41.46**	33.25**	32.26**	45.53***	23.57**	23.02**
	(18.80)	(15.42)	(15.70)	(13.33)	(11.68)	(11.67)
Constant	626.3***	444.3***	440.4***	473.7***	261.7***	270.2***
	(70.43)	(73.70)	(73.42)	(41.74)	(59.96)	(59.68)
Observations	522	522	522	649	649	649
Number of groups	19	19	19	19	19	19

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Figure captions to be inserted

Figure 1: Correlations between CLASS domain-level scores

Figure 2: Domain-level CLASS scores and teacher value-added

Figure 3: Teacher CLASS Scores by Value-Added Estimates

Figure 4: Association between Teacher Value-Added and CLASS Scores by Reason for Selection

Figure 5: Association between Teacher Value-Added and CLASS Scores by School Type