Article:

Darling, JC orcid.org/0000-0001-7176-5957, Bardgett, RJM and Homer, MS orcid.org/0000-0002-1161-5938 (2017) Can children acting as simulated patients contribute to scoring of student performance in an OSCE? Medical Teacher, 39 (4). pp. 389-394. ISSN 0142-159X

https://doi.org/10.1080/0142159X.2017.1291926

Can children acting as simulated patients contribute to scoring of student performance in an OSCE?

Jonathan C Darling[1], Rebecca JM Bardgett[2], Matthew Homer[1]

1 Leeds Institute of Medical Education, School of Medicine, University of Leeds

2 Bradford Teaching Hospitals NHS Foundation Trust

Running Head: Can children score students in an OSCE?

Total word count:	4103
Word count (excluding abstract, acknowledgements, notes on contributors, declaration of interest, figure legends and references):	3147

Corresponding author:
Dr Jonathan C Darling
Senior Lecturer in Paediatrics and Child Health
Division of Women's and Children's Health
Level 9 Worsley Building, University of Leeds, Clarendon Way
Leeds
LS2 9NL
Tel: +44 113 3431926
Email: j.c.darling@leeds.ac.uk

Practice points

- The child voice is often not heard in an OSCE, but children's views should be heard in all aspects of life that affect them
- Paediatricians cannot accurately predict children’s scores
- Simulated patient scores of children aged 8-10 years can make a meaningful contribution to scoring of student performance

Abstract

Background

The voice of the child is increasingly recognised as important, as summed up in the Department of Health report 'No decision about me, without me'. In medical education however, the child's voice is little heard: often in a paediatric OSCE, the examiner assigns a mark for the child.

Aim

To explore whether children can contribute meaningfully to summative scoring of student performance.

Methods

We studied this in two phases: first we compared child scores (CS) to examiner predictions of the child scores (EPCS), and other simulated patient (SP) scores within a single exam. Then we looked at CS over a further 4 exams.

Results

The Pearson correlation between CS and EPCS was 0.40 ($p < 0.001$), therefore EPCS accounted for 16% of variation in CS. Across 4 exams, the mean CS was higher than the mean adult SP score: exploratory factor analysis indicated that both may be measuring the same characteristic. Cronbach's alphas (0.66 to 0.76) did not significantly increase when SP scores (including CS) were removed.

Conclusion

Although there was some correlation between CS and EPCS, paediatricians could not accurately predict CS. We conclude that the child's voice can and should be heard within the OSCE marking process.

Introduction

The voice of the child and young person is increasingly valued, particularly in education, health and research. Initiatives such as the annual 'Takeover Day' where children and young people take over the roles of adults hundreds of organisations across the UK (Office of the Children's Commissioner, 2014) illustrate Article 12 of the UN Convention on the Rights of the Child (United Nations, 1989), which states that children should be able to have a say in matters that affect them. In the education sphere, children interview prospective teachers(Henry, 2005; Mansell, 2010; Williams, 2010). In health, the UK Department of Health has explicitly applied its catchphrase "No decision about me, without me" to children as well as adults, as part of the NHS drive towards shared decision-making in healthcare (Anonymous, 2012). The Royal College of Paediatrics and Child Health (RCPCH) has fleshed out this broad policy direction in a "how-to" guide that includes a case study on how their Youth Advisory Panel was involved in the appointment of their new CEO (Wood, Turner, & Straw, 2010). Children are increasingly involved in shaping research that may impact upon their lives, for example through the Young Person's Advisory Group set up by the NIHR (Young Person's Advisory Group, 2016).

Within medical education, our School of Medicine has been considering how best to involve children in undergraduate medical OSCE exams. We have previously reported the feasibility of involving a class of primary school children in a large-scale OSCE (Darling & Bardgett, 2013), and the perspective of the child in this setting (Bardgett, Darling, Webster, & Kime, 2015). However, there is very little published literature on whether children acting as simulated patients can contribute to scoring of student performance for use in pass/fail decision-making.

Our OSCE and involvement of children

In the Leeds Medical School 4th year OSCE (Bardgett et al., 2015; Darling & Bardgett, 2013), medical students undergo an end of year OSCE covering all Year 4 specialities: paediatrics; O&G/sexual health; psychiatry; chronic and continuing care (including primary care, palliative care and oncology); and acute and critical care (including emergency medicine and anaesthetics). School children aged 8-10 years from a local primary school are invited to participate as patients and undergo a limited 'system examination' (for example, lower limb neurological exam) by several students, as part of an OSCE circuit of 10 different stations. Each station lasts for 8 minutes, with 1-minute reading/preparation time between each station. In this OSCE setting, marks from other (adult) SPs (who are trained actors) are included in the scoring of students' performance, and have been shown to increase reliability of the exam (Homer & Pell, 2009). However, prior to this study, children were not asked to award marks. Since then, we considered that children's perceptions of their experience when being examined by a medical student were important and could form a valid component of the marking system.

Aims and phases of the study

The aim of the study was to explore whether child marking of student performance is valid and has similar statistical characteristics to other SP scores.

We studied this in two phases: in Phase 1, we compared their scores to examiner predictions of their scores, and other simulated patient scores and other metrics within the same single exam. In Phase 2, we looked at children's scores over a further 4 exams, but without examiner predictions of how they would score.

Methods

Phase 1

In 2009, 28 primary school children aged 8-10 years participated in the Year 4 OSCE, for 262 students. The children underwent cranial nerve examination, and were asked to assign a mark out of 10 in response to the question '*If you had to see a doctor again, how happy would you be to see this one*?' We used a 10-point scale to obtain more detailed responses for later analysis. No specific descriptors were given for each point on the scale. Normally SPs use a 5-point scale with detailed generic descriptors. The paediatrician examiners were asked to independently predict the child's response to the same question. In addition, they marked the station as usual, and gave a global rating regarding the overall individual student performance of the task. The global ratings for all 262 students were used to set the station pass-mark, which then contributed to the overall OSCE pass-mark. To investigate the relationship between the child score (CS) and other scores, the CS was correlated against:

- examiner prediction of child score (EPCS);
- sub-scores for organisation and communication (SSOC) within that station (on the basis that these components might well explain the child's perception of the interaction);
- examiner global rating (EGR), which is separate from the station marks, and used to set the pass-mark using borderline regression (Homer & Pell, 2009; Pell et al., 2010).

CS was then also correlated against summated SSOCs in all other stations, and with summated SP scores for the 10 stations with an adult SP, whose scores were analysed similarly.

Phase 2

From 2010 to 2014, children continued to participate in the OSCE exam in a similar way to Phase 1, although different body systems were examined within the paediatric stations of each exam (e.g. cranial nerve examination, upper or lower limb neurological examination, cardiovascular examination). For 3 of these OSCEs (2011 was omitted for logistical reasons), children were asked to give a response using a 5-point Likert scale to the same question as above, using the same response options as the adult SPs (0=strongly disagree, 1=disagree, 2=neutral, 3=agree, 4=strongly agree). We used a different question to that used for adult simulated patients because we judged that the question usually posed to the adult SPs was too adult-orientated for children to answer meaningfully. (The adult SP question was "*I felt that the student showed respect for me and responded to my concerns and questions in an appropriate and professional manner*.") This mark was included in the overall station mark. Children were briefed prior to each exam by one of the authors (JCD). In 2015, the briefing was extended to include discussion of example videos of an examination being performed, to more closely align their preparation to that provided for other SPs. Parental consent was obtained for children to participate in the exam. Children received a small amount of pocket money (£5) and a certificate as a 'thank-you' for their contribution. In 2013, the total number of stations in the 'main' OSCE was reduced from 18 to 16 as a part of a move to sequential testing (Pell et al., 2013). The second sequence of the exam contained a further 10 stations, but involved a smaller number of students, and data from this part of the exam has not been included in this paper. However, this did not affect the number of children participating as SPs, because the station remained within the main exam.

In Phase 2 we compared the child score for each OSCE with the mean SP scores from other stations with an SP, and conducted exploratory factor analysis to determine whether the CS and SP scores were likely to be measuring the same student characteristic. We then determined whether removal of any SP score (including the CS) had an impact on the Cronbach's alpha for the exam, as a way of estimating contribution (or not) of that score to the overall reliability of the exam.

Results

Phase 1

CS and EPCS were highly positively skewed, but the latter were less skewed (Skewness -1.22 and -0.67 respectively) (Figure 1). The CS and EPCS both ranged between 2 and 10. The Pearson correlation between CS and EPCS was 0.40 (n=217, p <0.001), therefore EPCS accounted for 16% of variation in CS. Overall there was some agreement between CS and EPCS (Figure 2). Adding EGR as a co-variate to the EPCS did not make any significant difference to prediction of CS. However, there was a slight tendency for CS to be higher in stations where EGR was lower, i.e. children tended to score higher than predicted when the examiner's overall impression of the student's performance was less good.

Within the station, there was little evidence of correlation between CS and SSOC, or with EGR. This is in contrast to other stations, where SP scores were highly correlated with these measures.

Across the whole OSCE, CS did significantly correlate (r=0.20-0.26) with summated SSOC, and summated EGRs. However, there was no significant correlation with summated SP scores.

Phase 2

Mean CS was higher than mean SP score for other stations, although the difference was reduced in the 2015 exam (Figure 3). Exploratory factor analysis showed a single main factor in most of the exams, (although in 2014 there could be two main factors), with CS mostly loading on to this main factor . This indicates that across most of the exams, the SP scores (including the child's) were possibly measuring the same trait or characteristic in student performance. Cronbach's alpha ranged from 0.66 to 0.76, with no station mean SP score having a detrimental impact on this (i.e. increasing the Cronbach's alpha if removed from the analysis), except in 2014 when removal of the CS resulted in a small increase in Cronbach's alpha from 0.66 to 0.67. After the extended briefing to the children in 2015, there is tentative evidence that their scores became more closely aligned to the other SP scores (Figure 3) – in essence, the effect size for the difference between child and other SP scores is smaller in 2015.

Discussion

Overall we found reasonable agreement between child scores (CS), and the examiner prediction of the child score (EPCS). However, children's scores tended to be more generous when examiners predicted lower-middle scores, and to a smaller extent when the examiner global rating was lower. Child and examiner prediction of child score (EPCS) were both high-scoring (Figure 1), indicating that the interaction with the child was generally conducted well, as attested by examiner observation and child experience. This would be expected from senior medical students who have all had 6 weeks of paediatric clinical placements, and whose exam preparation no doubt takes account of the likelihood of a child examination station featuring in the OSCE. They *should* be able to complete the task competently, and put the child at ease in the process, in comparison with the less predictable nature of many other OSCE stations. Marks for the child examination station do tend to be higher, and with less spread, than for other SP OSCE stations.

Previous research in our medical school looking at adult SP contribution to OSCE marking (Homer & Pell, 2009) concluded that although the picture is complex, including these scores is statistically defensible, but is partly a question of philosophical approach. The implied question is how much do we want to reflect the (simulated) patient viewpoint in our marking, and if so, what weight do we give it? This earlier work showed that the SP mark, although positively correlated with the total mark for the station, often only explains less than 10% of the variability in that mark, perhaps implying that the SP mark is measuring a different aspect of student performance.

In the current study, examiners could only predict 16% of the variation in CS. It is not surprising that paediatric examiners were unable to fully predict how the child felt about each student, since the latter perception is the child's subjective judgment. In addition, examiners may be influenced by other aspects of student performance, as highlighted by the station mark-scheme. We therefore cannot assume that examiners can adequately express the child's voice, or act as their proxy to award marks for student interaction with the child in the exam context, as is common practice in paediatric exams.

The CS was not significantly influenced by the overall competence of the candidate, as measured by the EGR within the station. This reassures us that children were not unduly influenced by task-specific competencies, which ideally should not influence their marking (but is probably difficult to separate, even for adult SPs). For example, if a student did not test power or reflexes correctly, this should not be reflected in the child score (unless they cause discomfort or confusion in the process). It should be noted that for other SP stations involving adults (for example, breaking bad news), it may be more difficult to separate the patient-interaction element from the other parts of the task. This may explain why other SP scores correlated with organization and communication ratings (SSOCs), and examiner global ratings (EGRs), but child scores did not. An alternative explanation is that the Child Score is measuring a different facet of student performance. However, over the whole 2009 exam, CS did correlate with summated SSOCs and summated EGRs, suggesting that to some extent children are able to pick out the globally less competent students. Students who struggle to put a child at ease may be those who struggle more generally – perhaps this is partly to do with confidence, since it is widely accepted in paediatric practice that children easily detect the underconfident clinician. It would be interesting to explore in more detail which domains of student performance influence children's scores, and to map these to overall outcomes for the examination. This would inform weighting of the CS in relation to the check-list.

Children gave higher marks than other SPs participating in our exams, although this difference was less evident in 2015 (Figure 3). It should be noted that other SP stations are mainly focused around consultation, history taking, explanation, and management, and there are few stations where the focus is an SP or real patient being clinically examined, as happens in the child station. It is arguable that these are quite different station types, and may therefore measure different candidate attributes. It may be quite appropriate for the pattern of child scores to differ from those of other SPs. For example, the skills required to show empathy and put a patient at ease during examination of lower limb neurology might be different and possibly less demanding than those needed to sensitively break bad news, and adult SPs in the former would be expected to mark more generously than those in the latter. The difference in 2015 may relate to an improved style of pre-exam briefing for the participating children, including use of example videos demonstrating clinical examination skills, which more closely aligns with briefings given to adult SPs.

Our data across 5 exams indicates that child marks do not detract from the general contribution of SP marks to the reliability of the exam, and may add to it. The exploratory factor analysis indicates that the child scores seem to be measuring the same student trait or characteristic across most of our exams.

The main limitation of this work is that when we compare child and adult SP marks we may not be comparing like with like because of the different station tasks, as noted above. In other words any child-marking effect is confounded by station (task). Further, the use of a different question for the child rating score compared to that used by adult SPs (because the adult question did not seem child-friendly) may mean we are measuring a different attributes to that measured by the adult SP question. The ideal design for this work would be a series of adult and child SPs all undergoing different body systems clinical examinations, and answering the same child-friendly rating question for the same patient experience. This would allow analysis of a range of SP scores across similar tasks and between child and adult SPs. However, of course, this is not practical within our exam process.

There is very little published on whether children of this age can meaningfully contribute to student marks in an OSCE. Crossley et al.(Crossley & Davies, 2005; Crossley et al., 2005) concluded that children's scores, although individually meaningful, are not reliable enough for summative assessment, although this was in an outpatient clinic setting (quite different to an OSCE), and response rate was only 58%. Davies et al. (Davies et al., 2012) developed a validated questionnaire for children 8 years and over to evaluate emergency care episodes (again for real clinical practice, not OSCE exams). Howells et al. reported on use of a Paediatric Consultation Assessment Tool to assess communication in three-way paediatric consultations (Howells et al., 2010). This is a fairly complex form with 7 scores on a 7 point scale for 19 sub-items, and in out-patient clinics showed similar reliability for parents and children (coefficients of 0.70 and 0.66 respectively), These were longer consultations (about 15 minutes) involving history and examination. The actual age of children completing the forms is not given.

Conclusions

There was reasonable correlation between child score and examiner prediction of child score, (r=0.4), although children were more likely to assign higher ratings. Paediatricians could not accurately predict child scores. Within the station, child scores did not significantly correlate with sub-scores for organisation and communication, or with examiner global ratings, in contrast to adult SP scores. However, child scores did correlate with these sub-scores across the whole exam. Over several exams, children tended to give higher scores than adult SPs, and use of their scores did not reduce the reliability of exams (except for a marginal reduction of 0.01 in 2014). Our findings could be due to the different types of OSCE station for child and adult SPs, the use of a different rating question, or because the child score measures a different facet of student performance. Overall this data, and our previous papers, show that the child's voice can and should be heard within the formal OSCE marking process.

Figures and Tables

Figure 1
Bar chart of child scores (CS) and examiner prediction of child scores (EPCS) for 2009 (Phase 1)

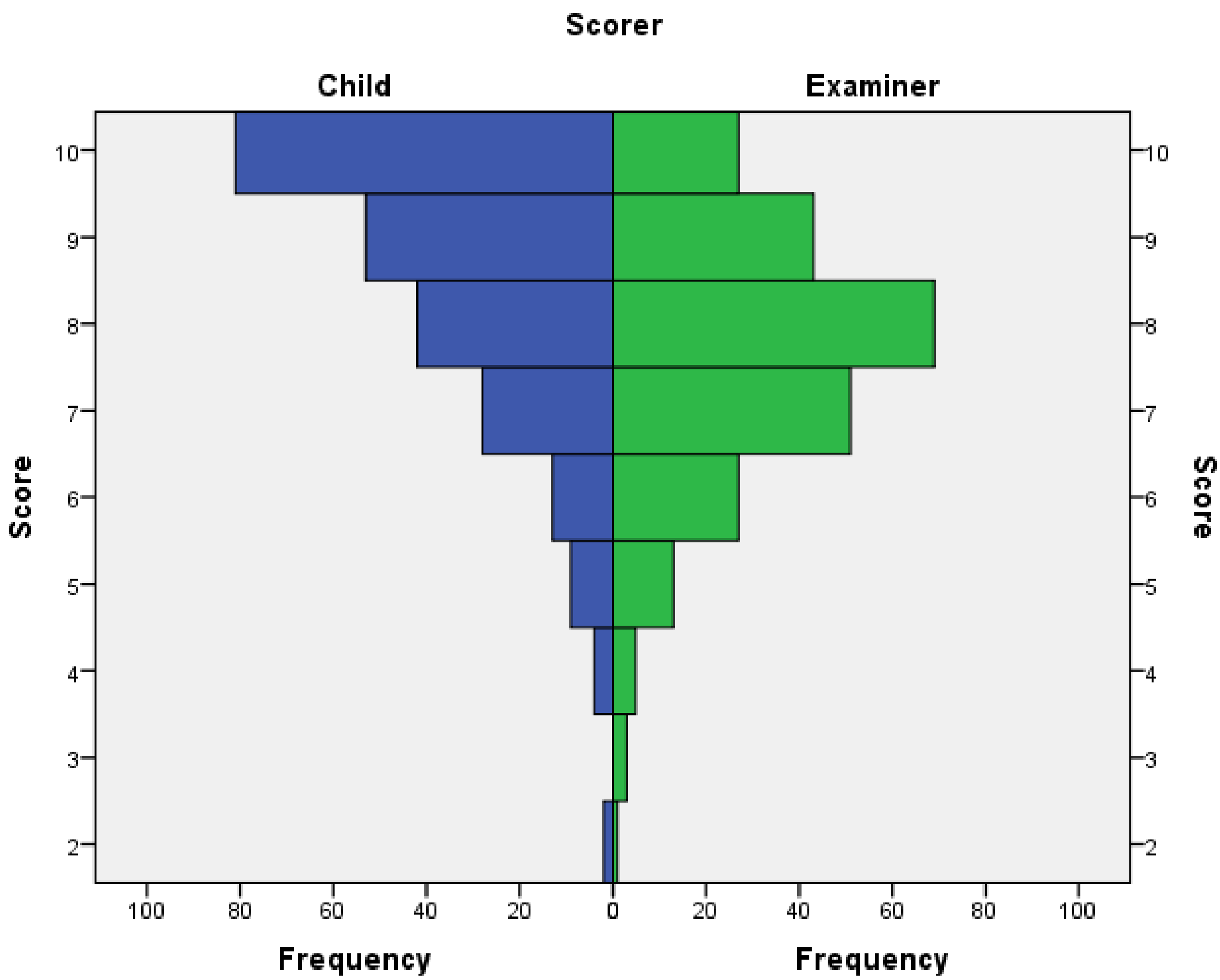
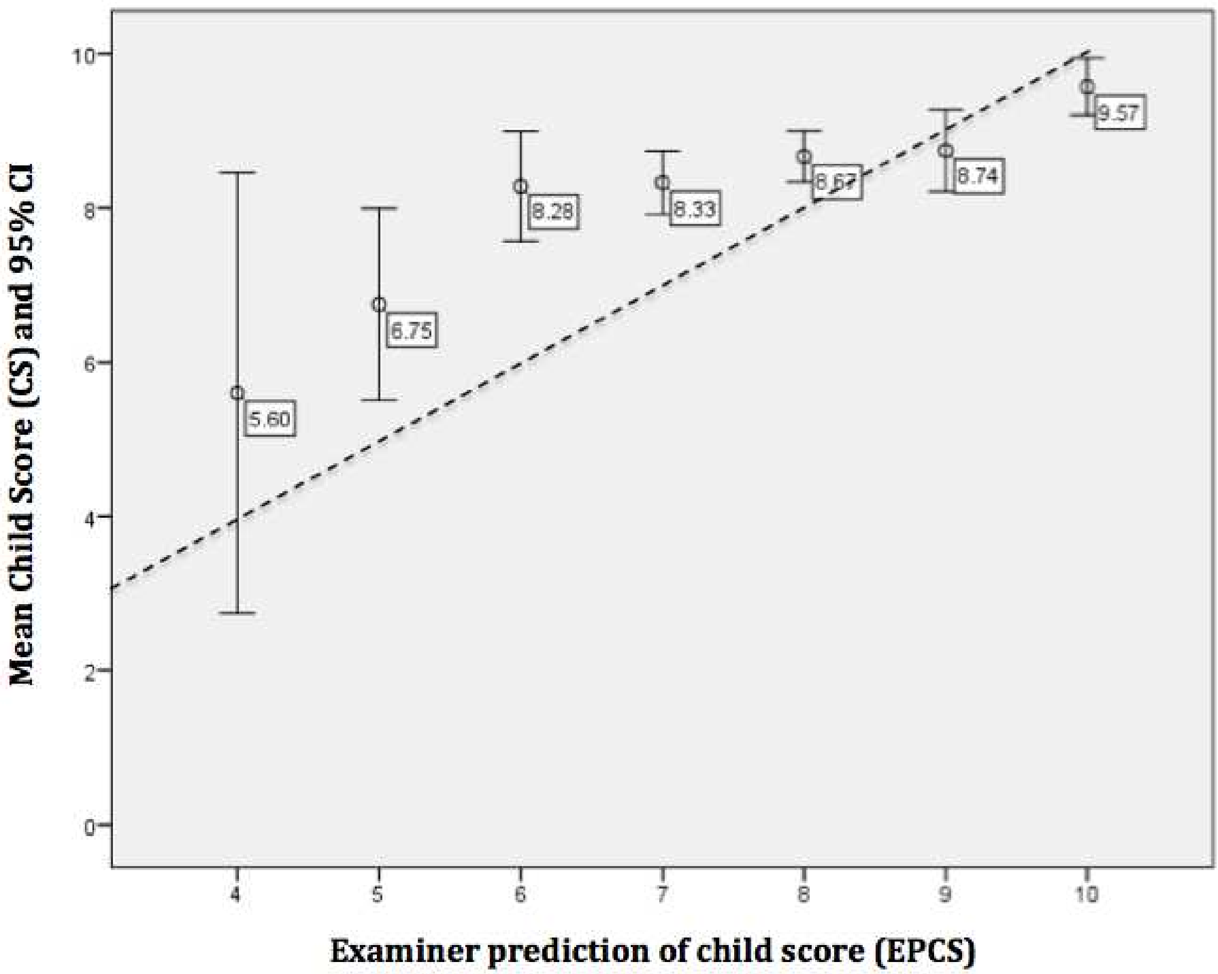
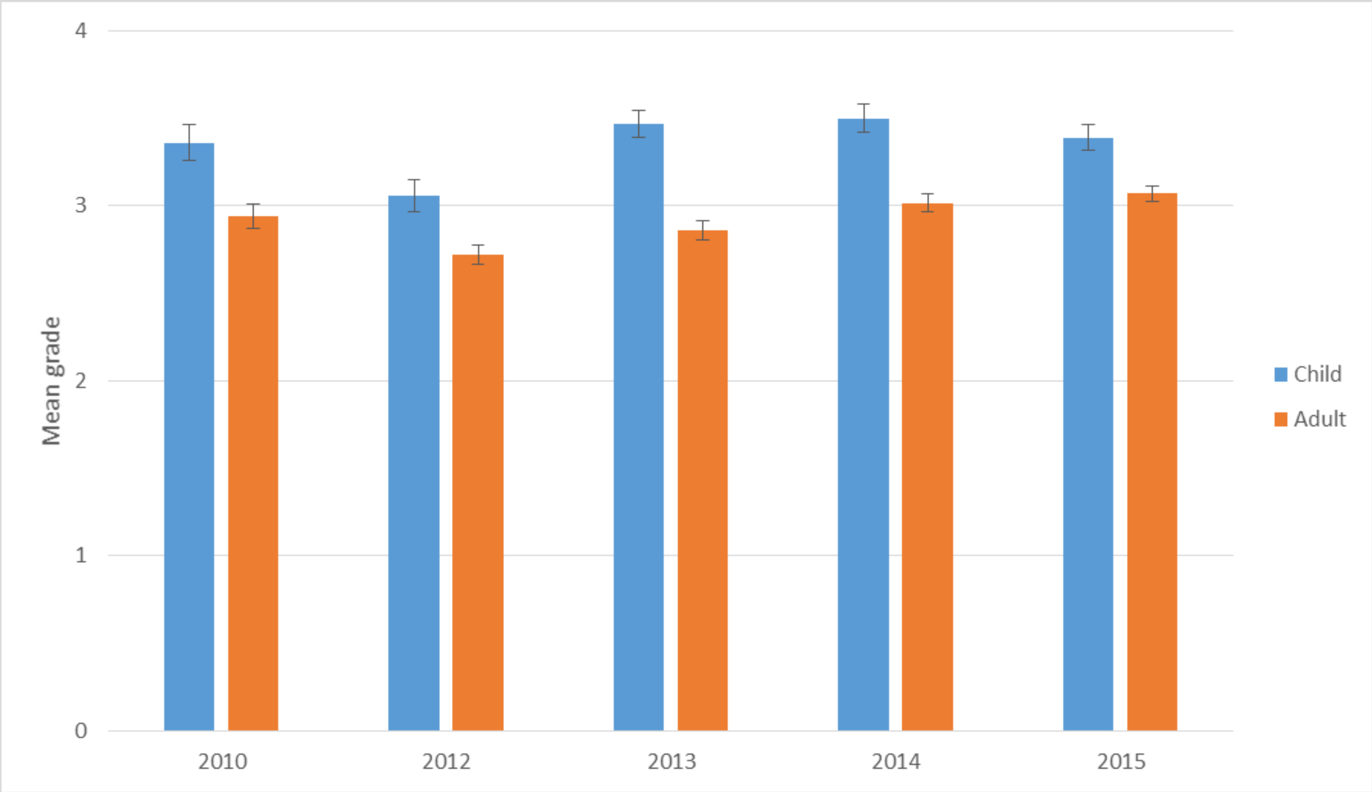


Figure 2
Error bar of mean child score (CS) for each level of Examiner Prediction of Child Score (EPCS) (Phase 1)



Note that error bars are longer at lower examiner scores because there are typically fewer cases. We do not show the mean child scores at EPCS of 2 and 3 (n=1 and 0 respectively). The dashed line indicates perfect agreement. See Figure 1 for further data, including number of data points for each score.

Figure 3
Mean child score (CS) and mean of other Simulated Patient (SP) scores for 2010-2015 (Phase 2)



p values for the differences between child and adult SPs were all statistically significant differences at the 5% level.

Note: no data for 2011.

Ethical approval

Phase 1 of the study was approved by the local medical education research ethics committee. No approval was necessary for Phase 2, since this is reporting our standard exam process.

Acknowledgements

We would like to thank the staff, children and parents of the primary school for their help with our exam. We would also like to thank the Year 4 examination team, who have gone out of their way to ensure a positive experience for all children involved.

Declaration of interest

The authors report no declarations of interest.

Notes on Contributors

JONATHAN C DARLING, MBCHB MRCP(UK) FRCPCH MD FHEA, is Senior Lecturer and Honorary Consultant in Paediatrics and Child Health at the University of Leeds and Leeds Teaching Hospitals NHS Trust. His research interests include clinical assessment in paediatrics and standard setting.

REBECCA J M BARDGETT, MBChB DTM&H MRCP(UK) MRCPCH, is Consultant Paediatrician at Bradford Teaching Hospitals NHS Foundation Trust. Her research interests include clinical assessment in paediatrics.

MATTHEW HOMER, BSc MSc PhD CStat, is an Associate Professor, working in both the Schools of Medicine and Education. His medical education research focuses on psychometrics and quality in assessment, particularly related to OSCEs and knowledge tests.

References

Anonymous. (2012). Liberating the NHS: No decision about me, without me. Government Response (pp. 1–42). London: Department of Health. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/216980/Liberating-the-NHS-No-decision-about-me-without-me-Government-response.pdf (Accessed 19/8/16).

Bardgett, R. J. M., Darling, J. C., Webster, E., & Kime, N. (2016). What makes a good children's doctor? Exploring the child perspective in the OSCE setting. *Medical Teacher*, 38, 471-475. http://doi.org/10.3109/0142159X.2015.1060301

Crossley, J., & Davies, H. (2005). Doctors' consultations with children and their parents: a model of competencies, outcomes and confounding influences. *Medical Education*, *39*(8), 807–819. http://doi.org/10.1111/j.1365-2929.2005.02231.x

Crossley, J., Eiser, C., & Davies, H. A. (2005). Children and their parents assessing the doctor-patient interaction: a rating system for doctors' communication skills. *Medical Education*, *39*(8), 820–828. http://doi.org/10.1111/j.1365-2929.2005.02230.x

Darling, J. C., & Bardgett, R. J. M. (2013). Primary school children in a large-scale OSCE: Recipe for disaster or formula for success? *Medical Teacher*, *35*(10), 858–861. http://doi.org/10.3109/0142159X.2013.806790

Davies, F., Vohra, J., & Hopwood, B. (2012). *The Development of a Patient Reported Experience Measure for Paediatrics Patients (0-16 years) in Urgent and Emergency Care: Research Report* (pp. 1–106). London: Royal College of Paediatrics and Child Health.

Henry, J. (2005, July 19). Pupils should help pick their own teachers. *The Daily Telegraph*. London. Retrieved from http://www.telegraph.co.uk/news/uknews/3321116/Pupils-should-help-pick-their-own-teachers.html (accessed 19/8/16).

Homer, M., & Pell, G. (2009). The impact of the inclusion of simulated patient ratings on the reliability of OSCE assessments under the borderline regression method. *Medical Teacher*, *31*(5), 420–425. http://doi.org/10.1080/01421590802520949

Howells, R. J., Davies, H. A., Silverman, J. D., Archer, J. C., & Mellon, A. F. (2010). Assessment of doctors' consultation skills in the paediatric setting: the Paediatric Consultation Assessment Tool. *Archives of Disease in Childhood*, *95*(5), 323–329. http://doi.org/10.1136/adc.2008.146191

Mansell, W. (2010, March 25). Pupil power: The children who chose their own head. *The Independent*. Retrieved from http://www.independent.co.uk/news/education/schools/pupil-power-the-children-who-chose-their-own-head-1926576.html (accessed 19/8/16).

Office of the Children's Commissioner. (2014). *Children's Commissioner's Takeover Day 2014 Evaluation Report* (pp. 1–54). London.

Pell, G., Fuller, R., Homer, M., & Roberts, T. (2013). Advancing the objective structured clinical examination: sequential testing in theory and practice. *Medical Education*, *47*(6), 569–577. http://doi.org/10.1111/medu.12136

Pell, G., Fuller, R., Homer, M., Roberts, T., International Association for Medical Education. (2010). How to measure the quality of the OSCE: A review of metrics - AMEE guide no. 49. *Medical Teacher*. Taylor & Francis. http://doi.org/10.3109/0142159X.2010.507716

United Nations (1989). *Convention on the Rights of the Child*. United Nations. Retrieved from http://www.ohchr.org/Documents/ProfessionalInterest/crc.pdf (last accessed 19/8/16).

Williams, R. (2010, April 3). Children asking 'frivolous' questions in teachers' job interviews. *The Guardian*, pp. 1–2. Retrieved from http://www.theguardian.com/education/2010/apr/03/children-job-interviews-questions (last accessed 19/8/16).

Wood, D., Turner, G., & Straw, F. (2010). *Not just a phase: A guide to the Participation of Children and Young People in Health Services* (pp. 1–56). London: Royal College of Paediatrics and Child Health. Retrieved from http://www.rcpch.ac.uk/system/files/protected/page/RCPCH_Not_Just_a_Phase_0.pdf (last accessed 19/8/16).

Young Person's Advisory Group. (2016). Young Person's Advisory Group. Retrieved April 8, 2016, from http://www.crn.nihr.ac.uk/children/pcpie/young-persons-advisory-group/ (last accessed 19/8/16).