OPEN ACCESS

Development of the Huddle Observation Tool for structured case management discussions to improve situation awareness on inpatient clinical wards

Julian Edbrooke-Childs,[1] Jacqueline Hayes,[2] Evelyn Sharples,[1] Dawid Gondek,[1] Emily Stapley,[1] Nick Sevdalis,[3] Peter Lachman,[4,5] Jessica Deighton[1]

► Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/bmjqs-2017-006513).

[1]Evidence Based Practice Unit, University College London and the Anna Freud Centre, London, UK
[2]Department of Psychology, University of Roehampton, London, UK
[3]Department of Health Service & Population Research, King's College London, London, UK
[4]International Society for Quality in Healthcare (ISQua), Dublin, Ireland
[5]National Clinical Lead SAFE, Royal College of Paediatrics and Child Health, London, UK

Correspondence to
Dr Julian Edbrooke-Childs, Evidence Based Practice Unit, University College London and the Anna Freud Centre, London NW3 5SU, UK; ebpu@annafreud.org

Received 19 January 2017
Revised 9 August 2017
Accepted 17 August 2017
Published Online First 19 September 2017

To cite: Edbrooke-Childs J, Hayes J, Sharples E, *et al*. *BMJ Qual Saf* 2018;**27**:365–372.

ABSTRACT

Background 'Situation Awareness For Everyone' (SAFE) was a 3-year project which aimed to improve situation awareness in clinical teams in order to detect potential deterioration and other potential risks to children on hospital wards. The key intervention was the 'huddle', a structured case management discussion which is central to facilitating situation awareness. This study aimed to develop an observational assessment tool to assess the team processes occurring during huddles, including the effectiveness of the huddle.

Methods A cross-sectional observational design was used to psychometrically develop the 'Huddle Observation Tool' (HOT) over three phases using standardised psychometric methodology. Huddles were observed across four NHS paediatric wards participating in SAFE by five researchers; two wards within specialist children hospitals and two within district general hospitals, with location, number of beds and length of stay considered to make the sample as heterogeneous as possible. Inter-rater reliability was calculated using the weighted kappa and intraclass correlation coefficient.

Results Inter-rater reliability was acceptable for the collaborative culture (weighted kappa=0.32, 95% CI 0.17 to 0.42), environment items (weighted kappa=0.78, 95% CI 0.52 to 1) and total score (intraclass correlation coefficient=0.87, 95% CI 0.68 to 0.95). It was lower for the structure and risk management items, suggesting that these were more variable in how observers rated them. However, agreement on the global score for huddles was acceptable.

Conclusion We developed an observational assessment tool to assess the team processes occurring during huddles, including the effectiveness of the huddle. Future research should examine whether observational evaluations of huddles are associated with other indicators of safety on clinical wards (eg, safety climate and incidents of patient harm), and whether scores on the HOT are associated with improved situation awareness and reductions in deterioration and adverse events in clinical settings, such as inpatient wards.

An area of concern for clinicians and researchers is evidence suggesting that England may have some of the highest levels of childhood mortality in Europe.[1] Some of the deaths could be due to failure to recognise the seriousness of the medical condition or to recognise deterioration in that condition.[2] Early detection of risk factors for patients, including deterioration on hospital wards, is key to improving patient outcomes.[3] There are multiple and complex causes of preventable morbidity and mortality in hospitalised patients, including unidentifiable and identifiable safety risks. Identifiable risks include delayed diagnosis of medical conditions, delayed recognition of deterioration, a lack of recognition of patient concerns and a lack of appropriate resources and staff.[4] Proposed solutions to address these identifiable risks, such as early warning system scores, are often restricted by fragmented approaches that fail to build capacity across hospitals, and focus predominantly on technical solutions as opposed to learning and cultural ones.[5 6] The integration of information is essential to achieving high levels of safety.

'Situation Awareness For Everyone' (SAFE) was a 3-year project aiming to redirect the clinical team's view of the patient and their disease or 'clinical gaze'.[7] In this process, a range of prospective indicators of risk or deterioration, including clinical indicators and staff concerns, are considered. The main intervention of the SAFE programme was the routine use of 'huddle' meetings on the wards. A huddle is an '*ad hoc meeting to re-establish*

situational awareness, reinforce plans already in place, and assess the need to adjust the plan'.[8] Situation awareness can be defined as '*the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future*'.[8] Increased situation awareness of potential risks in real time on inpatient wards may help reduce identifiable safety risks. For example, one study found that increasing situation awareness led to approximately 50% fewer unplanned transfers to higher levels of care.[9]

Huddles involve a suite of interventions to support and facilitate a ward culture of proactive rather than reactive care. The core tasks of a huddle are the identification of risks to patients, the development of a shared team understanding of patients who are at risk of deterioration and the making of plans to mitigate such risks. The teams would incorporate other tools such as early warning system scores and structured communication methods to aid these tasks. However, the way in which these tasks are achieved is also crucial in huddle theory. They must be brief, structured case management discussions of approximately 10–15 min duration in total, optimise staff engagement and focus on essential information only.[9 10]

As outlined in the literature,[9] a theory of change proposes that huddles enable collaborative and efficient information exchanges within ward teams, which fosters a shared clinical view of the current health state of patients. This promotes increased situation awareness, which leads to opportunities to identify actions that may be taken to mitigate risks and prevent deterioration of patients.[9] The introduction of huddles in the USA has demonstrated that huddles can lead to increased efficiency among staff members, improvements in the quality of information sharing, increased accountability, feelings of empowerment and a culture of collaboration and community. These work together to increase staff members' quality of awareness about patients and to facilitate staff members' capacity to enhance patient safety.[10]

AIM

The above evidence suggests that huddles may be effective at increasing levels of situation awareness in clinical settings, such as inpatient paediatric wards. However, how huddles are implemented in practice, and in different contexts, remains unclear. A review of the literature indicates that, to the best of our knowledge, there is no valid and reliable observational assessment available to apply to huddles. This is crucial when attempting to understand and improve how huddles take place and how they increase situation awareness, which is the primary desired outcome. Such a tool would help researchers and frontline staff to identify the essential ingredients of huddles and, in particular, it would help clinical staff to reflect on their practice to uncover areas for improvement. An observational instrument, like the Huddle Observation Tool (HOT), could help provide evidence for researchers and clinicians to determine whether huddles are implemented in line with best practice. Effectively, this could afford a reliable assessment of the fidelity with which the huddle intervention is applied (ie, whether huddles are delivered within wards as originally intended[11]). It would allow subsequent linkages between fidelity of implementation and effectiveness of the huddles in improving response to deterioration and enhancing the safety of patient care. The aim of the present research was to develop the HOT to capture important elements of huddles.

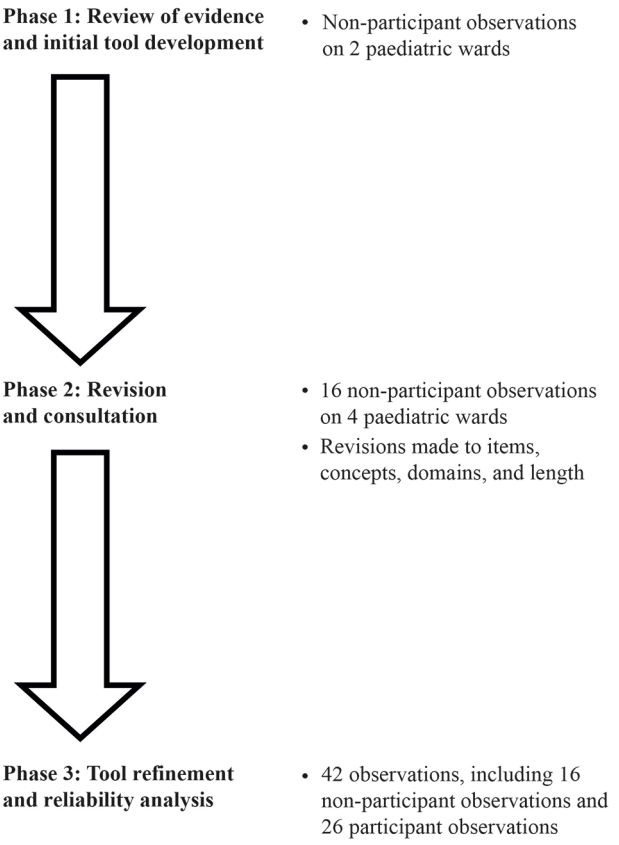


Figure 1 Summary of phases of development.

METHOD

A cross-sectional observational design adhering to relevant reporting guidelines[12] was used to develop the HOT through three standard psychometric instrument development phases (see figure 1). A favourable ethical opinion for the research was received from Dulwich Research Ethics Committee prior to data collection (reference: 14/LO/0875).

Phase 1: review of evidence and initial tool development

HOT version 01 (V01) was developed by using the following inductive process. A researcher (JH) reviewed the literature, with a particular focus on research generated by the original developers of

huddle methodology[9 10] and the design and use of observation tools in hospital settings.[13–16] An expert steering group was consulted, which included clinicians routinely working on paediatric wards (PL) and a senior researcher with over 10 years of expertise in developing clinical observation tools (NS).

Three researchers then conducted initial observations of huddles on two paediatric wards to identify the essential observable elements of huddles and field notes were recorded. These information sources were used together to generate an initial item pool for testing in the field for construct validity and reliability. Assessors were researchers with MSc level training in research methods. They were trained in non-participant observations of huddles in a 2-hour workshop before observations, with in-depth follow-up sessions as they reviewed these initial huddles to review their experience and any challenges. A training guide was also developed for HOT assessors (see online supplementary appendix A for instructions on how to conduct the observations and online supplementary appendix B for guidance on how to use the HOT) and observations were discussed and reviewed in team meetings. Any discrepancies in ratings of these initial huddles were discussed by the team until a consensus was achieved as to how huddles should be rated to ensure the assessors were ready to be deployed.

Phase 2: revision and consultation

HOT V01 was used to observe a total of n=16 huddles (9 morning huddles, 3 afternoon huddles, 3 evening huddles and 1 night huddle) across four paediatric wards participating in SAFE by four researchers over 2 months (January to February 2015).

Huddles lasted between 2 and 12 min and comprised between 2 and 16 members of staff, including doctors, nurses and allied health professionals. In addition to completing the tool, field notes were taken and huddles were audio-recorded. Sampling aimed for maximum heterogeneity of sites adopting the new initiative. Therefore, four paediatric units across England were selected, including two specialist children's hospitals and two district general hospitals, with the number of beds and length of stay considered to make the sample as varied as possible.

Substantive revisions were made to HOT V01 based on feedback from the observations. The key areas of feedback were that: (A) some words were ambiguous, (B) some items contained more than one concept ('all attendees have the opportunity to speak' vs 'some attendees are ignored and few speak'), (C) there was overlap between some of the domains assessed (eg, 'Culture' contained items on collaboration which were also contained in 'Coordination and cooperation'), and (D) the tool was too long, as huddles were generally not more than 5 min in duration. Through discussion with the expert steering group, HOT V02 was produced to more expediently capture the observable dimensions of situation awareness demonstrated through huddles. HOT V02 was viewed as more functional, with all quantitative data captured on one page.

The instrument was submitted for final refinement by the research team. Minor changes were made to HOT V02, predominantly consisting of simplifying the items and visual layout. HOT V03 was then used to rerate the 16 phase 2 audio-recorded huddles. The first author (JEC) and the last author (JD) rated the huddles blinded to each other, then discussed any rating discrepancies, and agreed the final ratings. No further changes to the instrument occurred as a result of this process.

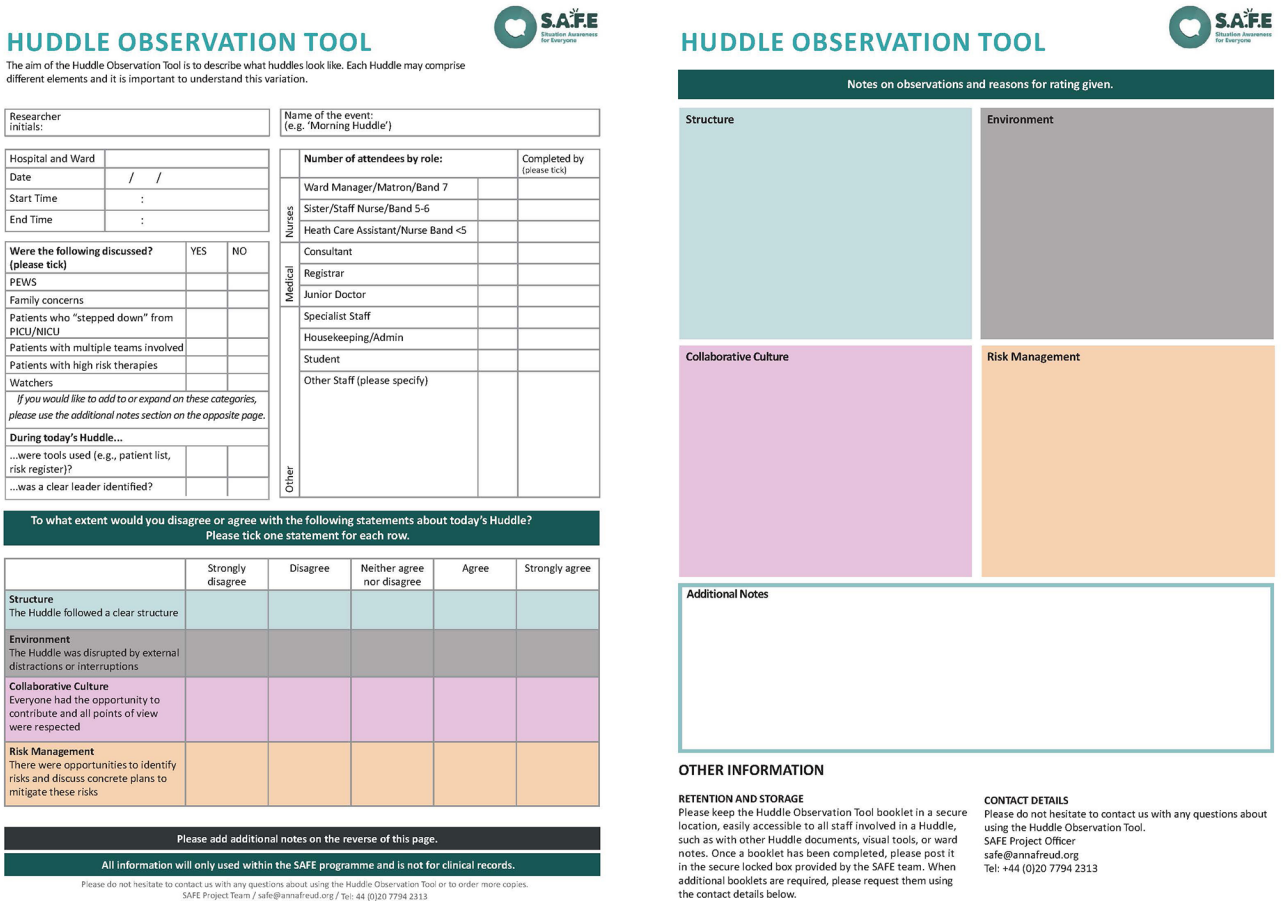
Phase 3: tool refinement and reliability analysis

Finally, for the main reliability analysis element of the study, huddles were observed across four paediatric wards participating in SAFE (as in phase 1) by three researchers over 16 months (February 2015 to May 2016, which was the duration of the overall evaluation[17]). This included non-participant and participant observations (the difference between the two was that in non-participant observations the observers were not members of the ward teams and did not interact at all with the ward teams during meetings, but in the participant observations they were active members of these teams). Huddles took place up to three times per day. The non-participant HOT V03 was used to observe a total of n=27 huddles (22 morning huddles, 5 afternoon huddles), of which n=16 (5 children's hospitals, 11 paediatric wards in district general hospitals) had two independent ratings for analysis. Huddles lasted between 2 and 11 min (median=4.5, IQR=3–7.25) and comprised between 3 and 20 members of staff (median=5, IQR=4–7), including doctors, nurses and allied health professionals (eg, physiotherapists).

Participant observation ratings were returned using the participant HOT V03 on n=30 huddles, of which 4 had missing data, resulting in a final sample of n=26 huddles. The overall data set for this phase was therefore 42 assessed team huddles with aggregated scores used when more than one observer rated a huddle. Huddles lasted between 2 and 15 min (median=10, IQR=5–10) and comprised between 3 and 20 members of staff (median=6.5, IQR=5.25–15), including doctors, nurses and allied health professionals (eg, physiotherapists).

As shown in figure 2, HOT V03 comprised four items rated on a 5-point scale from strongly disagree (0) to strongly agree (4) and was used with a global score[i] ranging from 0 to 16 each with free text response

[i]A global score was calculated; however, the unidimensionality of the instrument has not yet been examined. Future research, through the collection of a larger number of HOT assessments, should enable us to conduct a factor analysis of huddle observations, such that we can evaluate and establish the dimensional structure of the instrument. Larger scale data collection was beyond the scope of the current clinically focused project.



HUDDLE OBSERVATION TOOL

The aim of the Huddle Observation Tool is to describe what huddles look like. Each Huddle may comprise different elements and it is important to understand this variation.

Researcher initials:	Name of the event: (e.g. 'Morning Huddle')

Hospital and Ward	
Date	/ /
Start Time	:
End Time	:

Were the following discussed? (please tick)	YES	NO
PEWS		
Family concerns		
Patients who "stepped down" from PICU/NICU		
Patients with multiple teams involved		
Patients with high risk therapies		
Watchers		
If you would like to add to or expand on these categories, please use the additional notes section on the opposite page.		
During today's Huddle...		
...were tools used (e.g., patient list, risk register)?		
...was a clear leader identified?		

	Number of attendees by role:	Completed by (please tick)
Nurses	Ward Manager/Matron/Band 7	
Nurses	Sister/Staff Nurse/Band 5-6	
Nurses	Heath Care Assistant/Nurse Band <5	
Medical	Consultant	
Medical	Registrar	
Medical	Junior Doctor	
Other	Specialist Staff	
Other	Housekeeping/Admin	
Other	Student	
Other	Other Staff (please specify)	

To what extent would you disagree or agree with the following statements about today's Huddle? Please tick one statement for each row.

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
Structure The Huddle followed a clear structure					
Environment The Huddle was disrupted by external distractions or interruptions					
Collaborative Culture Everyone had the opportunity to contribute and all points of view were respected					
Risk Management There were opportunities to identify risks and discuss concrete plans to mitigate these risks					

Please add additional notes on the reverse of this page.

All information will only used within the SAFE programme and is not for clinical records.

Please do not hesitate to contact us with any questions about using the Huddle Observation Tool or to order more copies.
SAFE Project Team / safe@annafreud.org / Tel: +44 (0)20 7794 2313

HUDDLE OBSERVATION TOOL

Notes on observations and reasons for rating given.

Structure	Environment
Collaborative Culture	Risk Management

Additional Notes

OTHER INFORMATION

RETENTION AND STORAGE
Please keep the Huddle Observation Tool booklet in a secure location, easily accessible to all staff involved in a Huddle, such as with other Huddle documents, visual tools, or ward notes. Once a booklet has been completed, please post it in the secure locked box provided by the SAFE team. When additional booklets are required, please request them using the contact details below.

CONTACT DETAILS
Please do not hesitate to contact us with any questions about using the Huddle Observation Tool.
SAFE Project Officer
safe@annafreud.org
Tel: +44 (0)20 7794 2313

Figure 2 Huddle Observation Tool.

sections for notes. (1) 'Risk management' ('Were there opportunities to identify risks and come up with concrete plans for these risks?'), which was considered to clearly capture the key component of situation awareness. (2) 'Structure' ('Did the huddle have a clear structure?'), which was included to separate the format of a huddle from other concepts, such as collaborative culture and leadership. The role of a leader or coordinator was considered to be less salient than the organisation of the huddle (eg, a huddle could have a clear structure without a clear leader if all participants are aware of their role and turn in the discussion). (3) 'Collaborative culture' ('Did everyone have the opportunity to contribute and were all points of view respected?'), which was considered to be superordinate to the other items included in V01 (eg, a collaborative culture would necessitate all members, irrespective of their level of seniority, to be respected). (4) 'Environment' ('Was the huddle free from distractions?'), which was identified in the phase 1 observations as a salient component of huddles. The role of the leader was asked as a separate dichotomous question. The majority of huddles were rated as having a clear leader identified ($N_{non\text{-}participant\ \ observation}$=93%, n=9 missing; $N_{participant\ observation}$=23%, n=2 missing). Finally, the use of visual tools was asked as a separate dichotomous question. HOT V03 demonstrated face validity and was considered functional by all participant observers and by feedback from non-participant observers during site visits.

Analytic strategy

To assess the reliability of HOT V03, data were entered into and analysed using SPSS V.21.[18] For phase 2, inter-item consistency was used, which is suitable when assessing single-construct scales (ie, the quality and consistency of huddles as reflected by the global score). For phase 3, our aim was to assess the structural element of HOT V03. To this end, we examined the inter-rater reliability to assess whether or not different raters responded in a consistent manner using intraclass correlation coefficients (ICC) for the global score; an ICC ≥0.61 is considered acceptable for clinical feedback and ≥0.71 for research purposes.[15] For the inter-rater reliability for each item, we calculated the weighted kappa coefficient, taking into account the ordinal type of data for items in Likert scale. Ratings were available for 16 huddles; given the small number of data points and the original 5-point response scale, responses were recoded to three response options to enable analysis (ie, 'strongly agree' or 'agree' was recoded as 2, 'neither' as 1, and 'disagree' or 'strongly disagree' as 0). We also examined the relationships between individual items, and the individual items and the global score.

Table 1 Phase 2 inter-item correlations and item-total correlations of the HOT V03

	Huddle domain				
	Structure	Collaborative culture	Environment	Risk management	Global score
Structure	1.00				
Collaborative culture	0.19	1.00			
Environment	0.74**	0.55*	1.00		
Risk management	–0.15	0.47	–0.03	1.00	
Global score	0.73**	0.75**	0.85**	0.39	1.00
Mean (SD)	2.00 (1.41)	2.25 (1.00)	3.38 (1.02)	2.56 (0.93)	10.19 (3.04)

n=16 huddles.
*p<0.05.
**p<0.001.
HOT V03, Huddle Observation Tool version 03.

RESULTS

Phase 2: revision and consultation

Overall, the main changes were to simplify the tool (ie, clarifying domains, reducing the number and content of items, adding dichotomous (Yes vs No) questions) and changes to the layout. A participant observation version with training guide was also developed to mirror the non-participant version, which was used and discussed by staff at one of the hospital wards. A conference on SAFE open to staff from the 12 hospitals taking part in the project took place from 28 to 30 April 2015 during phase 2 (further data on conference attendance are not available). HOT V02 was presented to attendees and feedback recorded.

The descriptive statistics and results of the phase 2 reliability analysis are shown in table 1. There were two significant positive inter-item correlations between the structure and environment items, and collaborative culture and environment items. This suggests that more structured and collaborative huddles were conducted in environments with fewer interruptions. The correlation between the structure and environment items was 0.70, which is recommended when measures are designed to tap into the same underlying construct.[19] The smaller remaining inter-item correlations were not necessarily surprising as they were designed to tap into four different domains that may in fact be orthogonal. There were three significant positive item-total correlations between the structure, collaborative culture, and environment items, and the total score, and these were all above recommended values.[19] Huddles with higher levels of structure, collaborative culture and uninterrupted environments had higher overall global scores, suggesting that these elements were key to the quality and consistency of huddles.

Phase 3: prospective reliability analysis

The descriptive statistics and results of the phase 3 inter-rater reliability analysis are shown in table 2. The weighted kappa was only above the recommended value of 0.70 for the environment subscale. Still, the CIs did not overlap for the collaborative culture and risk management subscales. Although the CIs did overlap for the structure subscale, the majority of ratings were the same: 13 out of 16 ratings were rated as disagree or strongly disagree by both raters. Finally, the ICC was above the recommended value of 0.71 for the total score, suggesting that raters' scores were similar for this element.[15]

The descriptive statistics and results of the phase 3 internal consistency reliability analysis are shown in table 3. There were three significant positive inter-item correlations between structure and collaborative culture (large correlation), collaborative culture and risk management (moderate correlation), and risk management and environment (moderate correlation). These suggest that more structured huddles were associated with more collaborative cultures. There were also more opportunities to discuss risk management in huddles within more collaborative cultures and more opportunities to discuss risk management in huddles with fewer interruptions. However, only the correlation between structure and collaborative culture was

Table 2 Phase 3 non-participant descriptives and inter-rater reliability of the Huddle Observation Tool V03

	Huddle rating				
	Structure	Collaborative culture	Environment	Risk management	Global score
Mean (SD) rater 1	3.13 (0.62)	3.13 (0.89)	1.44 (1.46)	3.00 (0.37)	10.69 (2.41)
Mean (SD) rater 2	3.19 (0.66)	3.00 (1.21)	1.56 (1.46)	3.06 (0.77)	10.81 (2.79)
ICC (95% CI) or weighted kappa (95% CI)	0.45 (–0.15 to 1)	0.32 (0.17 to 0.42)	0.78 (0.52 to 1)	0.30 (–0.42 to 1)	0.87 (0.68 to 0.95)

n=16 huddles, which is less than the recommended number of 30.[22]
ICC, intraclass correlation coefficient.

Table 3 Phase 3 participant inter-item correlations and item-total correlations of the Huddle Observation Tool V03

	Huddle domain				
	Structure	Collaborative culture	Environment	Risk management	Global score
Structure	1.00				
Collaborative culture	0.76***	1.00			
Environment	0.29	0.36	1.00		
Risk management	0.45	0.40*	0.53**	1.00	
Global score	0.77***	0.77***	0.79***	0.74***	1.00
Mean (SD)	3.31 (0.97)	3.23 (0.71)	2.54 (1.36)	3.23 (0.65)	12.31 (2.85)

n=26 huddles.
*p<0.05.
**p<0.01.
***p<0.001.

above the recommended value of 0.70.[19] All four domains showed large significant positive item-total correlations,[20] and these were all above recommended values.[19]

DISCUSSION

The aim of the present research was to develop the HOT, which captures the essential ingredients of huddles. HOT V03 demonstrated face validity and was considered functional by participant and non-participant observers. In the phase 3 analyses, inter-rater reliability for the non-participant observations was acceptable for environment and the total score. It was low for collaborative culture, structure and risk management, although this may be partly explained by the small number of observations. Still, this suggests that collaborative culture, structure and risk management are more variable in how observers rate them but, nevertheless, agreement on the total score for huddles was acceptable. In the participant observations, the correlations suggest that more structured huddles were associated with more collaborative cultures, and that there were more opportunities to discuss risk management in huddles with more collaborative cultures, as well as in huddles with fewer interruptions. Further research should examine the direction of causality in these correlated huddle variables—for example, does a more collaborative culture lead to a better organised and structured huddle? Or does a clear huddle structure facilitate participants to achieve more collaboration? Qualitative research with participants of huddles will have a crucial role in indicating the most likely pathways through which these aspects of huddles influence each other, which is being captured in other parts of the SAFE evaluation.[17]

In addition, different huddles were rated by the two types of rater, and future research should examine whether these differences are explained by differences between huddles or between raters. Future research with a larger sample of observed huddles should continue to examine the reliability of the HOT; for instance, in terms of factor structure and internal consistency. Future research should also examine whether (A) data from participant and non-participant observations of huddles are associated with other indicators of safety on paediatric wards, such as safety climate and (B) whether huddles improve situation awareness and reduce preventable death on paediatric wards. These are important research questions that we are addressing in ongoing analysis from the SAFE evaluation programme.

In addition to future research, the HOT enables clinicians to view, record and reflect on—in a structured and precise manner—how a team communicates and identifies risks, whether this is part of a huddle or another type of case management discussion. This could provide useful information to help understand the processes around how a team currently minimises potential patient risks and subsequent harm. In turn, this may form the basis of continuous quality improvement of team and ward-level processes and patient care. HOT could be used as an anonymous feedback tool if given to all staff to rate their own huddles in order to inform continuous quality improvement of huddles.

Limitations should be considered when interpreting the findings of the present research. A small sample size of huddles was observed, particularly for the inter-rater reliability analysis, and the findings may not generalise to other huddles conducted across SAFE sites. Assessors took notes on the huddle in real time and completed the HOT ratings immediately after the huddle and before any discussions of the huddle between observers took place. Our training was very clear on the need not to calibrate or compare ratings between raters at the time of the data collection; anecdotal feedback from our raters confirms this is what they actually did on the wards. Although huddles were observed by researchers, the final ratings in the phase 1 analysis were based solely on audio recordings, meaning that valuable visual or non-verbal data may have been lost. Equally, however, a strength of this approach is that audio recordings enable the most accurate capture of the verbal-interactional features of the huddle in situ and do not rely solely on observer memory and note taking.[21] Self-selection bias may

apply, in that participant observation tools may have only been completed and returned by staff with more positive experiences of huddles. Future research should therefore recruit a random sample of participant observers.

Still, to the best of our knowledge, HOT is the first participant or non-participant observation tool for huddles. From a research perspective, HOT could be used to provide an objective measure of changes to huddles, situational awareness and collaborative culture over time, which is crucial when attempting to understand huddles and their role in increasing situation awareness. For this purpose, HOT was used over the course of the SAFE programme. From a clinical perspective, HOT could be used to assess the effectiveness of huddles as clinicians test and develop the best way to implement the huddle to improve teamwork and situation awareness, in line with best practice guidance in their ward. Although HOT was tested in paediatric wards, it may be useful for researchers and clinicians in reviewing and reflecting on huddles, and determining how they can most effectively be used to improve situation awareness in any clinical setting in which patient safety is a priority.

Thus, future research could also examine the application of HOT beyond huddles and paediatric wards. The implementation of huddles which are supported using HOT may be of relevance to adult wards, including intensive care and high dependency units. Inpatient mental health settings have an array of different safeguarding concerns and huddles may be a useful means of promoting situation awareness in these settings. Future research should also examine modified versions of HOT to capture other safety improvement interventions. For example, as part of the SAFE programme, wards have been implementing 'druggles', which are huddles specifically about patient medication, with the aim of minimising medication errors. We hope that the present research will help clinicians and researchers to be able to systematically analyse huddles and explore how they are implemented and, in turn, how implementation can be improved. The development of routine huddles and other approaches to reviewing safety is of particular interest, and the evaluation of this type of patient safety committee is too often forgotten.

Contributors JEC led the research and drafting of the paper under the supervision of JD, and JEC and JD rated the huddles and provided oversight of the research. JH conducted the literature review and developed the initial observation tool. ESh, DG and ESt conducted the observations of the huddles. NS and PL provided expert input on revising the observation tool. All authors contributed to the drafting of the paper.

Funding Situation Awareness For Everyone (SAFE) is a Health Foundation funded programme; both the implementation of SAFE and the evaluation were funded by the Health Foundation. This work was also supported by funding from WellChild, the funding was specifically to support evaluation work around perspectives of parents and young people and to support patient and parent involvement in the research. This programme of work and evaluation was also supported by the Royal College of Paediatrics and Child Health (RCPCH) which leads on the delivery of the programme. Sevdalis' research is supported by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care South London at King's College Hospital NHS Foundation Trust. Sevdalis is a member of King's Improvement Science, which is part of the NIHR CLAHRC South London and comprises a specialist team of improvement scientists and senior researchers based at King's College London. Its work is funded by King's Health Partners (Guy's and St Thomas' NHS Foundation Trust, King'sCollege Hospital NHS Foundation Trust, King's College London and South London and Maudsley NHS Foundation Trust), Guy's and St Thomas' Charity, the Maudsley Charity and the Health Foundation. Deighton was supported by the National Institute for Health Research (NIHR) Collaboration for Leaderships in Applied Health Research and Care (CLAHRC) North Thames at Bart's Health NHS Trust. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, the Department of Health, or RCPCH.

Competing interests None declared.

Ethics approval Dulwich Research Ethics Committee.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1 Wolfe I, Cass H, Thompson MJ, *et al*. Improving child health services in the UK: insights from Europe and their implications for the NHS reforms. *BMJ* 2011;342:d1277.
2 Wolfe IN, Macfarlane AL, Donkin AN, *et al*. *Why children die: death in infants, children and young people in the UK Part A*: Royal College of Paediatrics and Child Health, 2014.
3 van Galen LS, Struik PW, Driesen BE, *et al*. Delayed Recognition of Deterioration of Patients in General Wards Is Mostly Caused by Human Related Monitoring Failures: A Root Cause Analysis of Unplanned ICU Admissions. *PLoS One* 2016;11:e0161393.
4 Berwick D. *A promise to learn - a committment to act: Improving the safety of patient in England*. London: Department of Heath, 2013.
5 Batalden PB, Davidoff F. What is 'quality improvement' and how can it transform healthcare? *Qual Saf Health Care* 2007;16:2–3.
6 Taylor-Adams S, Vincent CA. *Human factors approaches to the analysis of serious incidents safety in medicine*: Elsevier, 2000.
7 Lachman P. Redefining the clinical gaze. *BMJ Qual Saf* 2013;22:888–90.
8 Situation awareness global assessment technique (SAGAT). *The National Aerospace and Electronics Conference (NAECON)*. New York: IEE, 1988.

9 Brady PW, Muething S, Kotagal U, *et al*. Improving situation awareness to reduce unrecognized clinical deterioration and serious safety events. *Pediatrics* 2013;131:e298–e308.
10 Goldenhar LM, Brady PW, Sutcliffe KM, *et al*. Huddling for high reliability and situation awareness. *BMJ Qual Saf* 2013;22:899–906.
11 Carroll C, Patterson M, Wood S, *et al*. A conceptual framework for implementation fidelity. *Implement Sci* 2007;2:40.
12 von Elm E, Altman DG, Egger M, *et al*. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ* 2007;335:806–8.
13 Flowerdew L, Gaunt A, Spedding J, *et al*. A multicentre observational study to evaluate a new tool to assess emergency physicians' non-technical skills. *Emerg Med J* 2013;30:437–43.
14 Nagpal K, Abboudi M, Fischler L, *et al*. Evaluation of postoperative handover using a tool to assess information transfer and teamwork. *Ann Surg* 2011;253:831–7.
15 Russ S, Hull L, Rout S, *et al*. Observational teamwork assessment for surgery. *Ann Surg* 2012;255:804–9.
16 Walker S, Brett S, McKay A, *et al*. Observational Skill-based Clincial Assessment tool for Resuscitation (OSCAR): Development and validation. *Resuscitation* 2011;82:835–44.
17 Deighton J, Edbrooke-Childs J, Stapley E, *et al*. Redirecting the clinical gaze on pediatric wards to improve safey outcomes and ensure Situation Awareness for Eveyone (SAFE): A multi-method realistic evaluation evaluation protocol. *BMJ Open* 2016;6:e014014.
18 *IBM SPSS Statistics for Window [program]. 21.0 version*. Armonk, NY: IBM, 2012.
19 Terwee CB, Bot SD, de Boer MR, *et al*. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60:34–42.
20 Cohen J. *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum, 1988.
21 Sacks H. *Lectures on conversation*. Oxford: Blackwell, 1992.
22 Terwee CB, Mokkink LB, Knol DL, *et al*. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res* 2012;21:651–7.