Inter-rater reliability in the Paediatric Observation Priority Score (POPS)

Authors:

Mrs Lisa Langton¹, Dr Adam Bonfield¹, Dr Damian Roland^{1,2,3}

Corresponding author: Dr Adam Bonfield, ab798@le.ac.uk, 07837717543

¹Department of Emergency Medicine, University Hospitals of Leicester, Leicester Royal Infirmary, Infirmary Square, Leicester, LE1 5WW, United Kingdom

²SAPPHIRE group, Health Services, University of Leicester, Leicester

³Paediatric Emergency Medicine Leicester Academic (PEMLA) group, Leicester Royal Infirmary

Word count: 2345

Abstract

Objective The primary objective of this study was to determine the level of inter-rater reliability between nursing staff for the Paediatric Observation Priority Score (POPS).

Design Retrospective observational study

Setting Single centre paediatric emergency department

Participants 12 participants from a convenience sample of 21 nursing staff.

Interventions Participants were shown video footage of three pre-recorded paediatric assessments and asked to record their own POPS for each child. The participants were blinded to the original, in person POPS. Further data were gathered in the form of a questionnaire to determine the level of training and experience the candidate had using the POPS score prior to undertaking this study.

Main outcome measures Inter-rater reliability amongst participants scoring of the POPS.

Results Overall Kappa value for case 1 was 0.74 (95% confidence interval 0.605 to 0.865), case 2 was 1 (perfect agreement) and case 3 was 0.66 (95% confidence interval 0.58 to 0.744).

Conclusion This study suggests there is good inter-rater reliability between different nurses use of POPS in assessing sick children in the Emergency Department.

Introduction

Triage scores are traditionally the method used in emergency departments (ED) to risk assess and manage the flow of patients entering the system. These are designed to identify sick patients and allocate a priority sequence. Triage categories do not absolutely quantify the severity of illness, or accurately predict the length of stay or eventual outcome(1).

The Paediatric Observation Priority Score (POPS) is a risk stratification tool designed to aid assessment of children presenting to the emergency and urgent care setting. POPS balances the need to identify the most unwell patients rapidly whilst avoiding poor sensitivity, which results in low thresholds for hospital admission, poor resource management and patient and family care(2). It also facilitates discharge of patients with minor and mild symptoms to primary care(3).

The POPS (figure 1) incorporates physiological parameters, such as heart rate, as well as more subjective observations, for example level of concern. Its derivation has previously been described(4) and it has undergone preliminary external validation(5). Physiological parameters within the normal range score a 0. Variation outside the norm scores 1 or 2 depending on degree of derangement. This gives a total score of 0 to 16.

POPS was designed to reduce cognitive load and aid decision making(2). There was a specific intent at its inception not to link POPS scores with actions, as opposed to a PEWS system, which is designed for this purpose. Given the wide range of outcomes in Emergency and Urgent care, and that the majority of children are likely to go home, a

prescriptive approach was felt to be unhelpful. The only local guidance we have recommended is that for very high scores (8+; less than 1% of attendances (6)) the child should be considered for the Emergency Room area (if they are not already there).

POPS was specifically designed for use in the emergency and urgent care setting by registered nurses, regardless of their level of clinical and/or paediatric experience. An initial training package involving group teaching and competency packs familiarises staff with the use of POPS.

The primary objective of this study is to establish if variation exists between different nurses use of the POPS assessment tool.

Methods

Video footage of three pre-recorded paediatric assessments were shown to the participants who were asked to record the POPS of each child. These videos had previously been piloted on experienced nursing staff not involved in the study, to determine face validity of the recordings.

Blinding was ensured between all of the participants and the in-person POPS.

Participants recorded data on standard documentation with access to the POPS matrix, as used in normal clinical practice. Further data were gathered in the form of a questionnaire to determine the level of training and experience the candidate

previously had using the POPS. No extra training was given to the candidates prior to this study.

Once data had been collected it was analysed manually using Microsoft excel 2007.

The data sets were checked for completeness and errors. The Fleiss Kappa statistic was utilised for the calculation of inter-rater reliability. This calculates proportion of agreement for each parameter, in addition to giving an overall kappa value between participants' case by case. The values were cross referenced using online calculator Recal 3(7).

Ethics approval was sought and approved by DeMontfort University Ethics committee.

Consent for video recording of patients was sought prior to their POPS being measured. Participants were recruited voluntarily, information leaflets stressing that competency was not being assessed in the research.

Results

A convenience sample of paediatric emergency department nursing staff, comprising 21 people, were invited to participate in this study. Of this, twelve nurses were consented to the study, each completing a POPS score for three cases, giving a total of 36 data sets.

Participants included a variety of nursing experience (UK equivalent band 5-band 7), with an interquartile range of 8 years 8 months qualified (median 9 years 11 months).

This sample is representative of the staff profile across the department. Table 1 gives a profile of the participants.

There were 3 video cases for the participants to score. Case one is described below.

The background and data to the other two cases can be found in appendix 1.

Table 2 gives an overview of the participants overall POPS for each video and a comparison with the POPS given to the child on admission.

The child subject for case 1 is an 8-week-old girl. She was brought into the ED by her mother with increased work of breathing. She had a past medical history of congenital heart disease. The Paediatric Advanced Nurse Practitioner (PANP) assessing her recorded the following observations:

Oxygen Saturations 70% (her normal is 70-85%);

Work of breathing, mild;

Alert;

Level of concern, low;

Past medical history of congenital heart disease;

Pulse 168;

Respiratory rate 55;

Temperature 36.5;

Table 3 shows the data sets of the POPS given by the participants for case 1. The data sets for case 2 and 3 can be found in appendix 1. The kappa value demonstrating proportion of agreement between the 12 respondents' can be seen at the bottom of

each observational score. The overall Kappa value for case one was 0.74 (95% confidence interval 0.605 to 0.865).

Data shown in appendix 1 demonstrates that the overall kappa value for case 2 was 1, representing perfect agreement, and the overall kappa value for case 3 was 0.66 (95% confidence interval 0.58 to 0.744).

Discussion

Inter-rater reliability

It has been suggested that a kappa value in the range of 0.6 to 0.8 demonstrates good inter-rater reliability(8). Therefore, analysis of the results of this study indicate inter-rater reliability between nurses for the POPS score to be 'good'. Case 1 and 3 featured unwell children, with abnormal physiological parameters. Kappa values for these cases were 0.74 and 0.66 respectively. Case 2 featured a child with observations within normal limits and clinical appearance of a well child. Perfect agreement was reached in this case with a Kappa value of 1.

Learning from the individual cases

Case 1:

Total POPS ranged from 6 to 10. One participant scored a 6 (observer 2) and another scored a 10 (observer 3). On analysis of these individual data sets the scoring appears acceptable within the context of the cases and the other scores, therefore these are not believed to be outlying data.

Oxygen saturations achieved the lowest proportion of agreement with a value of 0.333. A feature in case 1 is that the normal range of oxygen saturation levels for this child is 70%-85% due to a background of congenital heart disease. This is outside the normal range defined by the POPS matrix. Oxygen saturations varied throughout the video, resulting in documented values in the data sets ranging from 68%- 73%. There are differences between participants response to this. 4 nurses gave a score of zero indicating observation is within normal range, 2 staff gave a score of one indicating a mild deviation from normal, and 6 staff gave a score of 2 indicating significant derangement. It is important to note that, regardless of a background baseline oxygen saturation, the POPS score should be implemented based on the observations recorded. This demonstrates an area that could be targeted in educating staff members on complex patients presenting to the ED.

The observational parameter of gut feeling also demonstrated a lower level of agreement, with a value of 0.591. Defined parameters for scoring gut feeling are not stated within the POPS tool. This was an active decision by the tools designers to ensure that staff would likely over-score i.e junior nursing and medical staff would likely score 1 or 2 when a more experienced staff member may score 0. This would ensure patients remained safe and also alert senior staff to the relative competencies of those undertaking the observations.

All other parameters score highly on observer agreement. Work of breathing and respiratory rate have a proportion of agreement of 0.833, each with only one of the 12 participants scoring differently. However, previous work using a similar methodology to assess difficulty in breathing has demonstrated poor reliability between

clinicians(9). Our small sample size is therefore a limitation in drawing conclusions from this.

Perfect agreement was demonstrated in AVPU, other factors (cardiac history was recognised and scored a 2 in all cases), pulse and temperature. It is expected that perfect agreement would be seen in the temperature parameter, because this was given to participants in the film.

Case 2:

The PANP recorded all parameters within normal limits, giving an overall POPS of 0. All of the participants also recorded the POPS as 0, demonstrating complete agreement across the cohort. This result indicates that when a child is physiologically well, the POPS, regardless of experience, will identify these children as well. This is important from a triaging perspective as identifying well children early should correlate with shorter duration of admission and a reduction in resource use, ultimately increasing efficiency. This would be a useful area of future research to confirm whether this conclusion can be drawn.

Case 3:

The total POPS for case 3 range from 3 to 5. The range is close, with no obvious outlying values. This case presented a child with an oxygen saturation of 90%. The POPS matrix states that >95% should score a 0, 90-94% a 1 and <90% a 2. Two observers made an error, scoring an Oxygen saturation level of 90% as a 2. Observer 2 recorded oxygen saturations as 96%, giving a corresponding parameter score of 0,

even though the saturations never displayed this value. It is difficult to state whether this is participant error or a limitation of the video methodology.

The lowest proportion of agreement is in gut feeling, with a value of 0.561. All other parameters scored highly on observer agreement. 'Oxygen Saturations' and 'Other' produced a proportion of agreement of 0.651 and 0.682 respectively. The parameters AVPU, respiratory rate and pulse show near complete agreement, all with values of 0.833. Work of breathing and temperature have perfect agreement. This is expected in the temperature recording, as this observation is given to participants in the film.

However, work of breathing assessment is a subjective observation. The good level of agreement shown between nurses scoring of work of breathing in all cases is encouraging, although as stated previously, the small sample size limits the ability to draw conclusions regarding this.

Comparison with the literature

With reference to inter-observer studies looking at other types of Emergency Severity Index (ESI), this study demonstrates that the level of inter-rater reliability with the use of POPS is comparable. Table 4 presents a summary of the findings from studies identified in the literature.

Inter-rater reliability in the use of POPS lies within the same range as studies such as Jordi et al (2015) and rates higher than in the studies such as Allen et al (2015). An absolute value at which inter-rater reliability confers reliability does not exist. However, favourable comparison with other studies is an encouraging finding.

Previous research into Emergency Severity Index (ESI) inter-rater reliability has utilised written scenarios (10–16) or direct patient observation with simultaneous (12,15) or consecutive (17) assessments performed by the nurse and the researcher.

The justification for selecting video footage as a data collection instrument is to overcome the flaws identified in previous research. For example, it has been questioned how well subjective elements can be captured in written scenarios (18) and that changing clinical parameters between assessments may occur in direct observation (19). Whereas video footage can be accessed by the researcher and staff at convenience, without the requirement to intervene in the running of the emergency department or on the patient journey. Multiple cases can be viewed by numerous staff, enhancing the quantity of data collected. These aspects convey similar advantages to the written scenario but also gives the participant the experience of patient factors, such as pain, emotion and distress, which are recognised to influence clinical decision making (18). Furthermore, the condition of the patient remains unaffected by the passage of time, overcoming this problem when performing successive assessment as performed with direct patient observation.

Limitations

The validity of the findings of this study are limited by the small sample size. This was determined by access to a maximum of 21 participants. The small scale of the study restricts the ability to compare these finding with other larger studies within the literature. However, in the absence of definitive measures of inter-rater reliability, comparison with other studies is utilized as a frame of reference.

It has been suggested that staff from dedicated paediatric ED's had greater accuracy in triage practice compared to other generic facilities(12). As this study is conducted exclusively in a paediatric department, generalizability of the findings is limited to the specific population studied, and possibly similar departments with an equitable staff and patient profile.

The use of video footage has not been previously utilized for the purpose of inter-observer studies in this context, thus, reliability of this instrument has not yet been established. The size of this study and available time limits have not permitted this process to take place. Therefore, the reliability of video footage as a data collection tool remains untested. Furthermore, the skill and dexterity of the participant in using equipment to measure pulse, oxygen saturation levels and temperature are not tested. This must be considered a limitation of the data collection instrument. A higher inter-rater reliability may be seen when using video recordings compared to participants performing the physiological measurements themselves. Finally, a higher inter-rater reliability may be seen due to the videos being observed away from the clinical area, without the distractions and competing demands of a busy emergency department.

Conclusion

This study has produced evidence that there is good inter-rater reliability between different nurses use of POPS in assessing sick children in the Emergency Department. Inter-rater reliability is comparable with that for other ESI's. The addition of this evidence to the existing work done in validation of POPS supports the aspiration for

the continued use of POPS, and its uptake to an increasingly wider network of emergency care facilities.

What is already known on this topic:

1. Inter-rater reliability has previously been shown to be good for triage systems both adult and paediatric.
2. The Paediatric Observation Priority Score, an assessment system as opposed to a formal triage process, has been internally and externally validated in emergency care settings.
3. There are limited studies evaluating the reliability of assessment systems specifically designed for paediatric emergency care.

What this study adds:

1. This study demonstrates that inter-rater reliability of the POPS is 'good' in children with abnormal physiological parameters.
- 2.
3. This study promotes a new method for measuring inter-rater reliability in the use of video recorded assessments.

References

1. Challen K, Roland D. Early warning scores: a health warning. *Emerg Med J* [Internet]. 2016;33(11):812–7. Available from: <http://emj.bmj.com/lookup/doi/10.1136/emmermed-2014-204250>
2. Roland D, McCaffery K, Davies F. Scoring systems in paediatric emergency care: Panacea or paper exercise? *J Paediatr Child Health* [Internet]. 2016 [cited 2017 Sep 8];52(2):181–6. Available from: <http://onlinelibrary.wiley.com/doi/10.1111/jpc.13123/full>
3. Lambert V. A systematic literature review to support the development of a National Clinical Guideline – Paediatric Early Warning System (PEWS). Dublin; 2014.
4. Roland D, Lewis G, Fielding P, Hakim C. The paediatric observation priority score: a system to aid detection of serious illness and assist in safe discharge. *Open J Emerg Med* [Internet]. 2016 [cited 2017 Sep 8];4:38–44. Available from: https://www.researchgate.net/profile/Damian_Roland/publication/303976106_The_Paediatric_Observation_Priority_Score_A_System_to_Aid_Detection_of_Serious_Illness_and_Assist_in_Safe_Discharge/links/5762922208ae5df6d5b1de91.pdf
5. Roland D, Arshad F, Coats T, Davies F. Baseline characteristics of the paediatric observation priority score in emergency departments outside its centre of derivation. *Biomed Res Int*. 2017;2017:1–6.
6. Roland D, Davies F, Coats T. The Paediatric Observation Priority Score (POPS): Outcomes Of 24000 Patients. *Arch Dis Child* [Internet]. 2014;99(Suppl 2):A24.2-A24.
7. Freelon D. ReCal3: Reliability for 3+ Coders - dfreelon.org [Internet]. 2011 [cited 2017 Sep 8]. Available from: <http://dfreelon.org/utis/recalfront/recal3/>
8. Boslaugh S. *Statistics in a nutshell* [Internet]. second. O'Reilly Media; 2012 [cited 2017 Sep 8]. 569 p. Available from:

[https://books.google.co.uk/books?hl=en&lr=&id=Fc4fBgAAQBAJ&oi=fnd&pg=PR3&dq=BOSLAUGH,+S.+\(2013\)+Statistics+in+a+Nutshell.+2nd+Ed.+Sebastopol,+California:+O%27Reilly+&ots=fiAcfOnEPr&sig=Sq6PEtDWHrtCLLuHNmbxhxrQE80#v=onepage&q&f=false](https://books.google.co.uk/books?hl=en&lr=&id=Fc4fBgAAQBAJ&oi=fnd&pg=PR3&dq=BOSLAUGH,+S.+(2013)+Statistics+in+a+Nutshell.+2nd+Ed.+Sebastopol,+California:+O%27Reilly+&ots=fiAcfOnEPr&sig=Sq6PEtDWHrtCLLuHNmbxhxrQE80#v=onepage&q&f=false)

9. Bekhof J, Reimink R, Bartels I-M, Eggink H, Brand PLP. Large observer variation of clinical assessment of dyspnoeic wheezing children. Arch Dis Child [Internet]. 2015;100:649–53. Available from: <http://dx.doi.org/10.1136/>
10. Dallaire C, Poitras J, Aubin K, Lavoie A. Emergency department triage: do experienced nurses agree on triage scores? J Emerg Med [Internet]. 2012 [cited 2017 Sep 8];42(6):736–40. Available from: <http://www.sciencedirect.com/science/article/pii/S0736467911010122>
11. Jordi K, Grossmann F, Gaddis GM, Cignacco E, Denhaerynck K, Schwendimann RR, et al. Nurses' accuracy and self-perceived ability using the Emergency Severity Index triage tool: a cross-sectional study in four Swiss hospitals. Scand J Trauma Resusc Emerg Med [Internet]. 2015;23(1):62. Available from: <http://www.sjtrem.com/content/23/1/62>
12. Travers D, Waller A, Katznelson J. Reliability and validity of the emergency severity index for pediatric triage. Acad Emerg Med [Internet]. 2009 [cited 2017 Sep 8];16(9):843–9. Available from: <http://onlinelibrary.wiley.com/doi/10.1111/j.1553-2712.2009.00494.x/full>
13. Allen A, Spittal M, Nicolas C. Accuracy and interrater reliability of paediatric emergency department triage. Emerg Med Aust [Internet]. 2015 [cited 2017 Sep 8];27(5):447–52. Available from: <http://onlinelibrary.wiley.com/doi/10.1111/1742-6723.12455/full>
14. Ekins K, Morphet J. The accuracy and consistency of rural, remote and outpost triage nurse decision making in one Western Australia Country Health Service Region. Australas Emerg Nurs J [Internet]. 2015 Nov [cited 2017 Sep 8];18(4):227–33. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S1574626715000592>

15. van Veen M, Teunen-van der Walle VFM, Steyerberg EW, van Meurs AHJ, Ruige M, Strout TD, et al. Repeatability of the Manchester Triage System for children. *Emerg Med J* [Internet]. 2010;27(7):512–6. Available from: <http://emj.bmj.com/cgi/doi/10.1136/emj.2009.077750>
16. Durani Y, Brecher D, Walmsley D, Attia MW, Loiselle JM. The Emergency Severity Index Version 4: reliability in pediatric patients. *Pediatr Emerg Care* [Internet]. 2009 Nov [cited 2017 Sep 8];25(11):751–3. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19938298>
17. Gravel J, Gouin S, Goldman R, Osmond M. The Canadian Triage and Acuity Scale for children: a prospective multicenter evaluation. *Ann Emerg Med* [Internet]. 2012 [cited 2017 Sep 8];60(1):71–7. Available from: <http://www.sciencedirect.com/science/article/pii/S0196064411018889>
18. Gerrish K, Lacey A. The research process in nursing [Internet]. John Wiley & Sons; 2010 [cited 2017 Sep 8]. Available from: [https://books.google.co.uk/books?hl=en&lr=&id=_PL-ON921VMC&oi=fnd&pg=PR12&dq=GERRISH,+K.+and+LACEY,+A.+\(2010\)+The+Research+Process+in+Nursing.+6th+Ed.+Chichester,+West+Sussex:+Wiley-Blackwell.&ots=MvIU61-opZ&sig=JUpKh9rSf_bMa24oZdQVs6WX_9E](https://books.google.co.uk/books?hl=en&lr=&id=_PL-ON921VMC&oi=fnd&pg=PR12&dq=GERRISH,+K.+and+LACEY,+A.+(2010)+The+Research+Process+in+Nursing.+6th+Ed.+Chichester,+West+Sussex:+Wiley-Blackwell.&ots=MvIU61-opZ&sig=JUpKh9rSf_bMa24oZdQVs6WX_9E)
19. Nadjm B, Jeffs B, Mtove G, Msuya W. Inter-observer variation in paediatric clinical signs between different grades of staff examining children admitted to hospital in Tanzania. *Trop Med Int Heal* [Internet]. 2008 [cited 2017 Sep 8];13(9):1213–9. Available from: <http://onlinelibrary.wiley.com/doi/10.1111/j.1365-3156.2008.02129.x/full>
20. Landis JR, Koch G. An Application of Hierarchical Kappa-type Statistics in the Assessment of Majority Agreement among Multiple Observers. *Int Biometric Soc.* 1977;33(2):363–74.

Tables & Figures

Paediatric Observation Priority Score (POPS) Chart

This chart is not a substitute for good clinical judgement and any concerns about the condition of a child should be brought to the attention of a senior nurse or doctor

Age	Score	2	1	0	1	2	Total Score	Priority
Any	Sats	<90%	90-94%	>95%	90-94%	<90%		
Any	Breathing	Stridor	Audible grunt or wheeze	No distress	Mild or Moderate Recession	Severe Recession	0-1	
Any	AVPU	Pain	Voice	Alert	Voice	Pain	2-3	
Any	Gut Feeling	High level concern	Low level concern	Well	Low level concern	Child looks unwell	4-7	
Any	Other	Oncology Patient	Significant PMH*		Significant PMH*	Congenital heart disease	8+	Immediate review
0-1	Pulse	<90	90 - 109	110 - 160	161 - 180	180+	Any child scoring above 8 should be considered for transfer to resus	
	RR	<25	25 - 29	30 - 40	41 - 60	50+		
	Temp	<35°	35 - 35.9°	36 - 37.5°	37.6 - 39°	39°+		
1-2	Pulse	<90	90 - 99	100 - 150	151 - 170	170+		
	RR	<20	20 - 24	25 - 35	36 - 50	50+		
	Temp	<35°	35 - 35.9°	36 - 37.9°	38.0 - 40°	40°+		
2-4	Pulse	<80	80 - 94	95 - 140	141 - 160	160+		
	RR	<20	20 - 24	25 - 30	31 - 40	40+		
	Temp	<35°	35 - 35.9°	36 - 37.9°	38.0 - 40°	40°+		
5-12	Pulse	<70	70 - 79	80 - 110	111 - 150	150+		
	RR	<15	15 - 19	20 - 25	26 - 40	40+		
	Temp	<35°	35 - 35.9°	36 - 37.9°	38.0 - 40°	40°+		
13-16	Pulse	<60	60 - 59	60 - 100	101 - 110	110+		
	RR	<12	12 - 14	15 - 20	21 - 25	25+		
	Temp	<35°	35 - 35.9°	36 - 37.9°	38.0 - 40°	40°+		

***Significant PMH includes:**

- Ex-premature
- Syndromic conditions
- Cardiac problems
- Asthma
- Diabetes
- Long term steroids
- All other chronic conditions

POPS is copyright (creative commons attribution non-commercial sharealike 4.0) Dr Damian Roland and Dr Ffion Davies 2010
This is version 1.3 August 2016

Figure 1: Paediatric Observation Priority Score (POPS) matrix. Note:

RR is respiratory rate, PMH is past medical history

Table 1: Profile of participants

Participant	NMC Registration	Current Banding	Qualified nursing experience (years)	POPS training
1	RN(Child)	5	>11yr	Not stated
2	RN(Adult) RN(Child)	7	>11yr	Online/video learning resource Group face to face
3	RN(Adult)SpQ	5	>11yr	Online/video learning resource Group face to face
4	RN(Child)	5	6-10yr	Individual one-to-one teaching
5	RN(Child)	5	>11yr	Group face to face Paper learning package Individual one-to-one teaching

6	RN(Child)	5	0-5yr	Paper learning package
7	RN(Adult)	5	0-5yr	Individual one-to-one teaching
8	RN(Child)	6	6-10yr	Paper learning package Individual one-to-one teaching
9	RN(Child)	5	0-5yr	Individual one-to-one teaching
10	RN(Child)	5	6-10yr	Online/video learning resource Group face to face Individual one-to-one teaching
11	RN(Child)	6	>11yr	Group face to face Paper learning package
12	RN(Adult)	5	0-5yr	Online/video learning resource Group face to face Individual one-to-one teaching

Table 2: Overview of video cases and the participants overall POPS

		Participant POPS score											
Video	In-person derived POPS	1	2	3	4	5	6	7	8	9	10	11	12
Case 1	7	7	6	10	9	8	9	9	7	7	9	9	9
Case 2	0	0	0	0	0	0	0	0	0	0	0	0	0
Case 3	4	4	3	5	5	4	4	3	4	4	4	4	5

Table 3: Data set of POPS for case 1

Observer	O ₂ Saturations (%)	Breathing	AVPU	Gut feeling	Other	Pulse (BPM)	Respiratory rate (BrPM)	Temperature (°C)	Total score
1	0	1	0	1	2	1	2	0	7
2	0	1	0	1	2	1	1	0	6
3	2	1	0	2	2	1	2	0	10
4	1	1	0	2	2	1	2	0	9
5	1	1	0	1	2	1	2	0	8
6	2	1	0	1	2	1	2	0	9
7	2	2	0	2	0	1	2	0	9
8	0	1	0	1	2	1	2	0	7
9	0	1	0	1	2	1	2	0	7
10	2	1	0	1	2	1	2	0	9
11	2	1	0	1	2	1	2	0	9
12	2	1	0	1	2	1	2	0	9
Kappa value	0.333	0.833	1	0.591	1	1	0.833	1	0.74

Table 4: Summary of literature pertaining to Emergency Severity Index scores (ESI's)

Author	ESI studied	Data collection strategy	Sample size	Result	Agreement (Landis and Koch)(20)
Bespoke ESI's developed for assessment of paediatric patients in ED					
<i>Dallaire et al (2010)</i>	<i>Canadian Triage & acuity Scale</i>	<i>Written scenarios</i>	<i>5 nurses each assessing 100 written scenarios</i>	<i>Kappa 0.44</i>	<i>Moderate agreement</i>
<i>Gravel et al (2012)</i>	<i>Canadian Triage & acuity Scale</i>	<i>Patient triaged, immediate second triage by research nurse</i>	<i>1,464 patients</i>	<i>Kappa 0.74</i>	<i>Good agreement</i>
<i>Jordi et al (2015)</i>	<i>ESI</i>	<i>Written scenarios</i>	<i>69 nurses each assessing 30 cases</i>	<i>Kappa 0.78</i>	<i>Good agreement</i>
<i>Travers et al (2009)</i>	<i>ESI</i>	<i>Part 1 Written scenarios</i>	<i>170 nurses each assessing 40 scenarios</i>	<i>Kappa 0.77</i>	<i>Good agreement</i>
		<i>Part 2 Simultaneous assessment by triage nurse and research nurse</i>	<i>498 patients</i>	<i>Kappa 0.57</i>	<i>Moderate agreement</i>

<i>Durani Y, Brecher D, Walmsley D, Attia MW, Loisel JM (2009)</i>	<i>Emergency severity Index Version 4</i>	<i>Written scenarios</i>	<i>17 nurses 16 physicians each assessing 20 written scenarios</i>	<i>Kappa 0.67</i>	<i>Good agreement</i>
<i>Van Veen M, Teunen-van der Walle VFM, Steyerberg EW, van Meurs AHJ, Ruige M, Strout TD, van der Lei J, Moll HA (2010)</i>	<i>Manchester triage System for children</i>	<i>Part 1 Written scenarios</i> <i>Part 2 Simultaneous assessment by triage nurse and research nurse</i>	<i>43 nurses each assessing 20 written scenarios</i> <i>198 patients each assessed by one out of 25 nurses plus 1 out of 2 research nurses</i>	<i>Kappa 0.83</i> <i>Kappa 0.65</i>	<i>Almost perfect agreement</i> <i>Good agreement</i>
<i>Allen AR, Spittal MJ, Nicolas C, Freed G (2015)</i>	<i>Australasian triage scale</i>	<i>Written scenarios</i>	<i>167 nurses each assessing 9 written scenarios</i>	<i>Kappa 0.27</i>	<i>Fair agreement</i>
<i>Ekins K, Morphet J (2015)</i>	<i>Australasian triage scale</i>	<i>Written scenarios</i>	<i>46 nurses each assessing 27 scenarios</i>	<i>Kappa 0.4</i>	<i>Moderate agreement</i>