The Vancouver sedative recovery scale for children: validation and reliability of scoring based on videotaped instruction

Andrew J. Macnab MD FRCPC*, Marc Levine PhD†, Ned Glick PhD‡, Norman Phillips MA§, Lark Susak BScN§, Mary Elliott RN¶

We originally developed and tested the Vancouver Sedative Recovery Scale (VSRS) to measure recovery from sedation following paediatric open heart surgery and reported excellent clinical inter-observer reliability. We now report a new study using videotape instruction of novice raters and videotaped case examples to determine whether the instruction produces adequate skill with the VSRS. Inter-rater reliability was assessed using videotapes of 16 children across a range of ages (six months to six years), and all levels of sedation (unresponsive to fully awake). Variably randomized subsets of six of the 16 test cases were randomly assigned to be rated by each of 16 video-instructed ICU staff volunteers, according to a balanced incomplete block design, such that every pair of raters assessed two children in common. The validity of the ratings from the video-instructed raters was assessed by comparison with "gold standard" scores from two experts who rated all 16 children clinically as the test cases were videotaped. The experts were in agreement themselves (intraclass correlation of 0.976). The correlation between the novice scores (average of six ratings per video) and the live clinical scores (average of two expert ratings) was 0.977 over the 16 test cases. On average, the mean expert rating was slightly higher, but the difference was negligible. (The differences between the mean ratings of the experts and novices for 13 of the 16 videos are very close to zero, while the other three differences, from technically less good videos, are two standard deviations away from zero). The VSRS, video instruction method and incomplete block design may be of use to other clinicians and investigators.

Pour mesurer le réveil et la sédation après la chirurgie à coeur ouvert chez l'enfant, nous avons développé et vérifié l'échelle de réveil et de sédation de Vancouver (VSRS) dont nous avons déjà communiqué la fiabilité. Nous pulions maintenant une nouvelle étude au cours de laquelle nous utilisons une vidéocassette pour former les appréciateurs et des exemples sur vidéocassette pour déterminer si cette méthode suffit pour produire une compétence suffisante avec le VSRS. La variabilité entre appréciateurs est évaluée avec des vidéocassettes de 16 enfants d'âges différents (de six mois à six ans) et à tous les degrés de sédation (de l'insensibilité au réveil complet). Des sous-catégories variables aléatoires de six des 16 observations expérimentales sont attribuées au hasard pour évaluation à chacun des 16 préposés volontaires déjà initiés à la vidéo à l'unité des soins intensifs, suivant un modèle de bloc équilibré incomplet de sorte que chaque paire d'appréciateurs doit évaluer deux enfants en commun. La validité des évaluations des appréciateurs formés par vidéo est établie par comparaison avec les scores de deux experts qui ont classifé par la clinique les 16 enfants au moment où on les enregistrait sur vidéo. Les experts sont d'accord entre eux (corrélation 0,976). La corrélation entre les appréciateurs novices (en moyenne six évaluations par vidéo) et les scores de références mesurés sur le vif (moyenne de deux évaluations d'experts) est de 0,977 pour les 16 observations. Les évaluations moyennes des experts sont légèrement plus élevées, mais cette différence est négligeable. Pour 13 des 16 vidéos, la différence entre les évaluations des experts et des novices est très près de zéro, alors que les trois autres différences pour des vidéos de moins bonne qualité technique s'éloignent de deux écarts-types du zéro. Le VSRS, la formation par vidéo et le modèle de bloc équilibré incomplet peuvent être utiles à d'autres cliniciens et chercheurs.

Key words
ANAESTHESIA: paediatric;
EDUCATION: videotape;
RECOVERY: assessment.

From the Departments of Paediatrics*, Pharmaceutical Sciences†, Health Care and Epidemiology‡, University of British Columbia, Susak Medical Research Associates§ and BC's Children's Hospital¶.

Address correspondence to: Dr. Andrew Macnab, Room 1M2, Intensive Care Unit, BC's Children's Hospital, 4480 Oak Street, Vancouver, BC V6H 3V4.

Accepted for publication 3rd June 1994.

Following any paediatric surgical or diagnostic procedure requiring general anaesthesia or sedation, it is important, in both research and clinical contexts, to assess the child's level of alertness after he or she has "recovered consciousness."

Studies have indicated that in children there can be a correlation between levels of consciousness and respiratory complications postoperatively.[1] Some children may be discharged before they are adequately recovered, and therefore still be at risk of post-anaesthetic complications such as aspiration, while others may be kept in the PAR for longer than is necessary. Since such care requires one-to-one nursing, this decision can be very costly in terms of staff hours. Using a valid scoring system allows earlier discharge of adult PAR patients than reliance on clinical criteria.[2] Paediatric versions of adult recovery scales exist,[3] but these scores include other measurements, particularly of respiratory adequacy, and do not focus on neurological recovery once the patient is awake. The VSRS is the first paediatric scoring system designed specifically for measuring levels of alertness following sedation/anaesthesia.

The Vancouver Sedative Recovery Scale (VSRS) (Table I) was developed and validated[4] to provide objective scoring of a child's level of alertness at any time during post-sedative recovery. The VSRS can assist staff to determine when it is safe and appropriate for the paediatric patient to leave the post-anaesthetic recovery room (PAR) or other areas such as endoscopy, the emergency room, or an intensive care unit (ICU).

As previously reported,[4] the VSRS was developed during ICU studies of various sedative protocols following open heart surgery in children. To develop the VSRS, a variety of indicators of levels of alertness among sedated children were identified, and the applicability and face validity of these indicators were determined. The original version of the VSRS consisted of 12 items that encompassed three categories of indicators: response; eye appearance and function; and body movement. The total possible score ranges from 0 to 22, with the higher scores indicating increasing levels of alertness. The VSRS was administered to 82 paediatric ICU and PAR patients, with each patient assessed simultaneously by at least two raters. Internal consistency as measured by Cronbach's alpha was excellent: 0.85. Inter-observer agreement or reliability as measured by intra-class correlation was also very high: 0.90. For individual items Cohen's kappa ranged from 0.65 to 0.89.

We were satisfied that the VSRS was a good attempt to quantify recovery from sedation in children. However, we believed that administration of the instrument could be improved with minor modifications. Therefore we clarified the descriptors for the four items scoring movement; the score for accommodation was simplified to "accommodates" or "does not accommodate" (deleting "no attempt to accommodate"), and the scoring of tremor was clarified by division into "major," "minor" or "absent tremor" (replacing "present" or "absent"). The total score for the scale remained the same and no new categories were added. Table I presents the current VSRS in full.

TABLE I The Vancouver Sedative Recovery Scale

		Score
Response		
A	(i) Awake/alert	4
	(ii) Awake/drowsy	3
	(iii) Asleep/easily aroused	2
	(iv) Asleep/difficult to arouse	1
	(v) Asleep/unable to arouse	0
Note: If child scores "0" on above, do not proceed		
B	(i) Responds fully to stimuli in an age-appropriate manner	2
	(ii) Delayed response to stimuli	1
	(iii) Absent response to stimuli	0
C	(i) "Alert" facial expression	1
	(ii) "Flat" facial expression	0
Eyes		
D	(i) Bright eyes	1
	(ii) Dull eyes; glazed	0
E	(i) Looks "at you"	1
	(ii) Looks "through you"	0
F	(i) Accommodates	1
	(ii) Does not accommodate*	0
G	(i) Recognition of stimulus	1
	(ii) Limited or no recognition of stimulus	0
H	(i) Purposeful and spontaneous eye movement	1
	(ii) Little or no spontaneous or purposeful eye movement	0
Movement		
I	(i) Spontaneous and varied central activity	4
	(ii) Spontaneous and varied peripheral activity	3
	(iii) Central activity in response to stimuli	2
	(iv) Peripheral activity in response to stimuli	1
	(v) No movement	0
J	(i) Absence of tremor or ataxia	2
	(ii) Minor ataxia or tremor	1
	(iii) Major ataxia or tremor†	0
K	(i) Coordinated spontaneous movement	2
	(ii) Weak/coarse spontaneous movement	1
	(iii) No purposeful spontaneous movement	0
L	(i) Shows age-appropriate manual dexterity	2
	(ii) Awkward or clumsy hand movement	1
	(iii) No fine hand movement	0

Note: Maximum score: 22 (fully awake); minimum score: 0 (unable to arouse).

*("No attempt to accommodate" deleted from original VSRS).

†Item added to original VSRS.

New work with the VSRS was undertaken in response to inquiries from prospective clinical users in various disciplines. We prepared a videotape to assist novice raters to learn the VSRS; and we videotaped a suite of 16 test cases to use for practice scoring. The present study, which uses a balanced incomplete block design to investigate the reliability of the video learning, examines the consistency of video test-case scoring among novice raters taught the VSRS using videotaped instruction; and calibrates their video-based scoring against "gold standard" clinical scores for the same test cases rated by two expert raters at the time of videotaping.

Videotaped instruction makes it possible to teach groups of raters to perform the VSRS with minimal disruption of staff and patients. Because the methodological innovations address statistical limitations that are common to other studies of reliability and are not specific to the VSRS, we anticipate both components of the study will be of use to other clinicians and investigators.

TABLE II Rating data (novice and expert) for 16 cases in incomplete block design

		CASES															
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
NOVICE RATERS	1				6				0	18		12		14			19
	2				5		4			13	13		0			16	
	3			20		13	1		0	15					15		
	4	9						22		19	18			14	14		
	5	11	0	21						19		10				19	
	6		1			13		22		16			0				22
	7	12	0		0		1	22	0								
	8		1	19	5	14					15			14			
	9				9	16		15				10			12	19	
	10	10		17	7								0		13		20
	11		0						0				0	13	16	22	
	12			20				20	0		12	10	0				
	13	6				15			0		17					19	22
	14	8				16	1					11	0	15			
	15			21			1	18						16		21	22
	16		0				1				18	15			10		21
EXPERT RATERS	1	10	0	21	8	18	4	16	0	18	16	10	0	16	12	19	20
	2	10	0	21	5	19	7	18	0	16	15	12	0	14	14	20	21

AWAKE SCALE

Methods

The study was approved by the University of British Columbia and BC's Children's Hospital Research Review Committees. Subjects for the VSRS assessments were recruited from paediatric patients between six months and six years of age recovering from general anaesthesia in the Post-Anaesthetic Recovery Room, and in the Intensive Care Unit. Children were excluded if they were handicapped, had major congenital anomalies, or if their family's first language was not English.

The study method, based on a balanced incomplete block design, involved comparison of ratings of videotaped test cases by a group of video-instructed raters with "gold standard" scores from two experts who rated children clinically as the test cases were videotaped.

We performed a series of "live" clinical VSRS assessments and videotaped each assessment. To obtain the greatest accuracy for the "live" assessments, we employed two research nurses who had nine months' experience in the clinical use of the VSRS; and for purposes of this study we considered these two expert raters to be our "gold standard." The assessments were performed in the clinical setting where excellent inter-observer reliability had been demonstrated in our previous study.[4] The two experts both scored each child independently (average assessment time: four minutes) and they were blinded to each other's scores.

Thirty "live" assessments with simultaneous video taping were performed in order to accumulate a pool of assessments covering a range of ages and levels of alertness. Had funding allowed, more children would have been videotaped to achieve optimal technical quality. The principal investigator then reviewed the tapes for acceptable photographic quality and appropriate content for the study. Sixteen taped assessments were selected on the basis of quality and representation of the range of children's ages (six months to six years of age) and levels of alertness (from unable to arouse to fully alert).

An instructional tape was produced to introduce novice raters (volunteers from Children's Hospital ICU staff) to the content and scoring procedure of the VSRS. The principal investigator oriented these raters to the purpose and design of the scale and showed them the instruction tape (45 min). The novices then practised rating two videotaped sample cases and received feedback from the PI. All test-case videos used different children to those in the sample cases and instructional videotape.

Each of the 16 video-instructed novice raters then scored a different subset set of six of the 16 videotaped test cases, selected according to the specifications of a balanced incomplete block design (Plan 11.27 from Cochran and Cox[5]), so that every pair of raters scored two children in common (see Table II). This design has an efficiency index of 0.89 relative to a completely crossed design (in which each of 16 nurses would rate each of 16 taped assessments). Ratings were obtained from each rater independently, in a structured setting.

Data analysis

Rater reliability was estimated by the intraclass correlation coefficient[6,7] obtained from the balanced incomplete block design described above.[5] An estimate of the

intraclass correlation coefficient as well as a one-sided confidence internal was computed as in Fleiss[8] using mean squares obtained from the MANOVA Procedure calculated with SYSTAT MGLH statistical software. Agreement between ratings of taped assessments and live ratings was assessed by the disattenuated correlation between the average rating for a taped assessment and the average of the two live ratings for the corresponding live assessment. The intraclass correlation for the live ratings was also calculated.

By choosing small subsets of cases according to the balanced "incomplete block" design,[5,8] the study retained power close to that which might be achieved if every rater scored every child. This reduced the number of video assessments that had to be scored by each rater.

Results

The validity of the video ratings obtained from the volunteer nurses was assessed by the agreement with the live ratings obtained from the two experts. The two experts were in agreement themselves as indicated by an intraclass correlation of 0.976. The average of six novice ratings per video was highly correlated with the average of two ratings (original live assessment) obtained from the two experts (correlation = 0.977), indicating considerable agreement. Table II contains each observer's rating of each case assessed. Figure 1 presents a scatter plot of these two sets of mean ratings for each of the sixteen videos. Figure 2 presents a bias plot to examine the agreement of the two methods of assessment over the range of scores.[9-12] Bias is indicated by the average difference between the two methods. In this case, the average rating given by the experts was slightly higher than that of the novices but the difference was negligible. The differences between the mean rating from six novices and the mean rating form two experts for 13 of the 16 videos are very close to zero, while the other three videos are approximately two standard deviations away from zero.

Discussion

The degree of recovery from analgesia or sedation has implications not only for the well-being of the child, but also for health care costs. Prolonged recovery from sedation, and its attendant morbidity, may result in the need for prolonged supervision and excessive duration of stay in PAR or intensive care. The accumulation of sedative drugs in a child's system may delay extubation because of inadequate respiratory effort even though the child no longer requires mechanical ventilation for other reasons. Even when a child is conscious enough for tracheal extubation, he or she may remain in a relatively sedated state with decreased spontaneous movement and ineffective cough, contributing to postoperative morbidity,

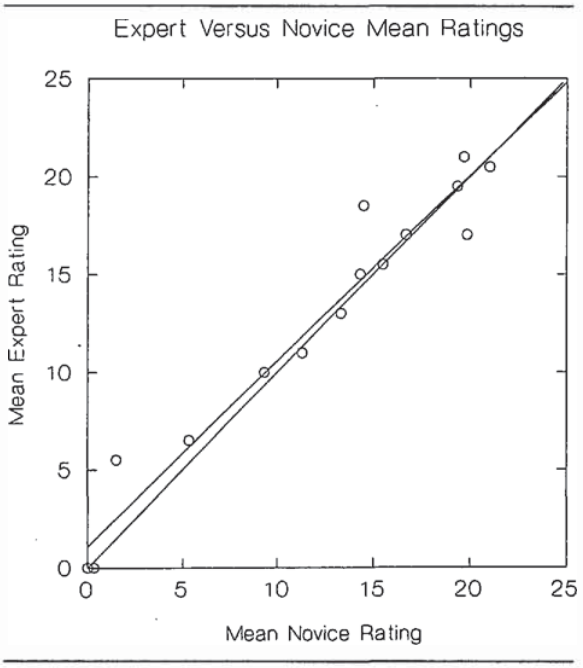

FIGURE 1 Scatter plot of expert and novice mean ratings for each of 16 videos (Vancouver Sedative Recovery Scale).

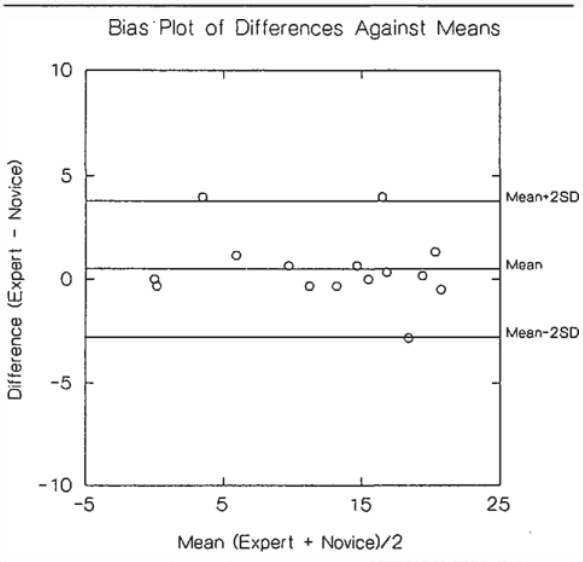

FIGURE 2 Bias plot to examine the agreement of two methods of assessment (live and video) over the range of scores (Vancouver Sedative Recovery Scale).

and increasing the incidence of atelectasis and reintubation. In order to be able to make an objective comparison of the duration and degree of sedation following the administration of different anaesthetic or sedative pro-

tocols, it is necessary to have a reliable assessment tool. To date no paediatric scale has been available.

There is little information in the literature about objective measurement of degree of sedation in children at various times after initial awakening has occurred. Although the assessment of adults has been addressed, studies have not been performed in the paediatric population, possibly because children are more difficult to assess and a "normal" response varies considerably with age.

Traditionally, assessment of recovery from sedation in adults is based on their orientation to time, place and person, and their ability to interact. The assessment process is more difficult in children, since infants are unable to communicate verbally and young children may be too apprehensive to communicate or cooperate.

Because we could not rely on communication for assessment of a child's degree of alertness, we identified a group of observations of response, eye appearance and function, and body movement that, in combination, would allow us to differentiate degrees of alertness. Any one observation alone would not be sufficient, particularly because of the broad range of ages and medical conditions of the children who are likely to be assessed using the scale. The combination of indicators and the range of scores is designed to be relatively unaffected by a child's physical or developmental abilities except in severe cases such as quadriplegia or blindness.

Although there are scales in the literature for the measurement of recovery from sedation (Observer's Assessment of Alertness/Sedation (OAA/S), Visual Analogue Scale (VAS)),[13] Post-Anaesthetic Discharge Scoring System (PADS),[3] they are either inappropriate for use in children, or are not sensitive to the degree of sedation once the child is conscious. Postoperative anaesthetic recovery scales are designed for the physical assessment of the patient in the recovery room, but measure levels of consciousness up to the point of awakening,[14-16] and the Glasgow Coma Score,[17] modified for use in children,[18] was devised to measure impairment of consciousness after head injury. These scales are not designed for the assessment of level of alertness after the child "regains consciousness."

The evaluation of inter-rater reliability is intended to apply to all potential raters in general. Any such inter-rater reliability statistic, usually the intraclass correlation coefficient, depends on estimating variability, not only among subjects, but also among raters. As inclusion of too few raters is an inherent shortcoming in many inter-rater reliability studies, our ideal study design would have included a large number of children to be assessed as subjects and a large number of raters present simultaneously at each assessment to score the child's alertness. The incomplete block design enabled 16 raters to assess 16 children effectively and had four major advantages. It avoided the presence of a large group of adults around the child's bed which would intimidate many children and could introduce bias by making some of them less cooperative. It avoided the excessive time commitment necessary for 16 patients to be visited separately, and all raters to score each child (raters' compliance likely would have been very low). It saved costs by maximizing the number of raters and minimizing their time commitment. Statistical power for the method was maintained. In addition, the video instruction method resulted in consistent instruction for all novice raters.

The video instruction method produced an acceptable level of performance of the VSRS. Although rating of 13 of the 16 test cases was excellent, three cases were less accurately rated. The scores for these three children were not in a particular range of the scale (Table II). On review, these video segments were not technically ideal (i.e., one did not proceed in the sequence that most raters would naturally follow during an assessment) and all three contained scenes which were either briefer than the comparable scenes in the other 13 cases, or not in perfect focus. Unfortunately, within the provision of the funding, video filming had to be limited. With optimum video quality, we anticipate that rater accuracy would have been higher than that achieved in this study. We feel that the outliers represent a deficiency in the video technique, rather than an inherent inaccuracy of the VSRS design. It is important to note that the nurse "experts" were very consistent in their clinical ratings of the three patients who were outliers.

It should be recognized that the group of raters from the Paediatric ICU were particularly experienced in subjectively assessing level of sedation in children, and more instruction might be necessary for staff from more general areas.

The VSRS should be useful for the objective assessment of level of sedation in paediatric patients requiring premedication, sedation, anaesthesia or postoperative care for emergency room procedures, diagnostic studies, oncology therapy, endoscopy, or surgery. The incomplete block design and the video instruction method may be of value to other clinicians and investigators in planning studies of a similar nature. Application of the design to the study of other scoring systems would seem logical.

Acknowledgements

We gratefully acknowledge the funding provided for this study by the BC Health Research Foundation and support from the BC Lung Association (Career Award – A.J. Macnab). We would also like to thank the ICU staff who participated in the study: Kerry Olihan (expert rater), Margit Bartha, Marry Bellis, Vena Camerizuli, Carol

Campbell, Wendy Crawford-Laurence, Janice de Boer, Lesley Forbes, Rosella Jefferson, Nelly Keutgen, Kathy Labuda, Roberta March, Lori McCreight, Sandy Norman, Mary Spencer, Frances Starr, and Corinne Williams.

References

1 *Parr SM, Robinson BJ, Glover PW, Galletly DC.* Level of consciousness on arrival in the recovery room and the development of early respiratory morbidity. Anaesth Intensive Care 1991; 19: 369–72.

2 *Wetchler BV.* Outpatient anaesthesia. What are the problems in the recovery room? Can J Anaesth 1991; 38: 890–4.

3 *Patel R, Rice LJ.* Special considerations in recovery of children from anesthesia. Int Anesthesiol Clin 1991; 29: 55–68.

4 *Macnab AJ, Levine M, Glick N, Susak L, Baker-Brown G.* A research tool for measurement of recovery from sedation: The Vancouver Sedative Recovery Scale. J Pediatr Surg 1991; 26: 1263–7.

5 *Cochran WG, Cox GM.* Experimental Designs, 2nd ed, New York: John Wiley & Sons, 1957.

6 *Shrout PE, Fleiss JL.* Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979; 86: 420–8.

7 *Bartko JJ.* The intraclass correlation coefficient as a measure of reliability. Psychol Rep 1966; 19: 3–11.

8 *Fleiss JL.* Balanced incomplete block designs for inter-rater reliability studies. Applied Psychological Assessment 1981; 5: 105–12.

9 *Bland JM, Altman DG.* Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986; 1: 307–10.

10 *Altman DG, Bland MJ.* Measurement in medicine: the analysis of method comparison studies. The Statistician 1983; 32: 307–17.

11 *Lee J, Koh D, Ong CN.* Statistical evaluation of agreement between two methods for measuring a quantitative variable. Comput Biol Med 1989; 19: 61–70.

12 *Cleveland WS.* The Elements of Graphing Data. Belmont CA: Brooks-Cole, 1985.

13 *Chernik DA, Gillings D, Laine H, et al.* Validity and reliability of the observer's assessment of alertness/sedation scale: study with intravenous midazolam. J Clin Psychopharmacol 1990; 10: 244–51.

14 *Steward DJ.* A simplified scoring system for the postoperative recovery room. Can Anaesth Soc J 1975; 22: 111–3.

15 *Aldrete JA, Kroulik D.* A postanesthetic recovery score. Anesth Analg 1970; 49: 924–33.

16 *Carignan G, Keéri-Szàntô M, Lavellée J-P.* Postanesthetic scoring system. Anesthesiology 1964; 25: 396–7.

17 *Teasdale G, Jennett B.* Assessment of coma and impaired consciousness. A practical scale. Lancet 1974; 2: 81–4.

18 *Simpson D, Reilly P.* Paediatric coma scale (Letter). Lancet 1982; 2: 450.