

Determining Alertness in Individuals with Profound Intellectual and Multiple Disabilities: The Reliability of an Observation List

Vera Munde, Carla Vlaskamp, Wied Ruijsenaars, and Han Nakken
University of Groningen, the Netherlands

Abstract: In the support of individuals with profound intellectual and multiple disabilities (PIMD), assessing the level of alertness is a recurring issue for parents and other direct support persons. Although observations show clear advantages above and beyond other assessment methods, there are problems related to this method as well. Subjectivity of interpretation and low reliability results have been described as the main problems. In the present study, our aim was to estimate the reliability of the Alertness Observation List (AOL) while, at the same time, minimizing the problems entailed in observations. We calculated both the inter-observer agreement and intra-observer agreement for 39 situations. Since the results exceeded the formulated 80%-criterion, we concluded that the AOL was a reliable instrument. However, the large range found in the results was striking. Moreover, observers with different information about the observed individuals came up with different reliability scores. To determine the value of observation of individuals with PIMD, it might well be necessary to judge the actual usefulness that the instrument has in clinical practice, besides the reliability of the results.

While stimulation to promote communication and learning is essential for the support of individuals with profound intellectual and multiple disabilities (PIMD) (Guess et al., 1993), direct support persons (DSPs) regularly wonder how to determine the “right moment” for starting such stimulating activities. By the same token, it is important for an activity to be started at the “right moment” so as to allow time for the stimuli that are presented to enter the consciousness of the individual with PIMD (Nelson, van Dijk, McDonnell, & Thompson, 2002). The “right moment” has also been described as “being focused on the environment” or as “being alert” (Munde, Vlaskamp, Ruijsenaars, & Nakken, 2009). The questions that arise, then, refer to a number of topics: How does an individual show that he or she is focused on the environment?

We would like to thank the children and teachers who agreed to participate in this study. Special thanks are owed to the observers for their time and effort. Correspondence concerning this article should be addressed to Vera Munde, Department of Special Education, University of Groningen, Grote Rozenstraat 38, 9712 TJ Groningen, the NETHERLANDS. E-mail: V.S.Munde@rug.nl

How can individual differences in alertness signals be interpreted? Can we determine an optimal moment during the day for stimulation of an individual with PIMD? There is an additional problem in that research shows that reduced levels of alertness and quick, irregular changes in alertness levels over time are common for individuals in the target group (Guess, Roberts, & Guy, 1999). These factors may even aggravate the problem of determining alertness reliably in individuals with PIMD.

Despite agreement about the importance of determining alertness for the support of individuals in the target group, it is not obvious *how* different alertness levels ought to be determined. Since individuals with PIMD do not express their needs by means of spoken language, self-report cannot be used (Vlaskamp, 2005). Similarly, physiological measurements often show unusual patterns and do not reveal the necessary information about the complex behavior of individuals in the target group (Mudford, Hogg, & Roberts, 1997). In contrast, most authors do agree that alertness can be described in terms of observable behavior. Consequently, most instruments used to investigate and determine alertness in individuals

with PIMD are based on observations (Munde et al., 2009).

Observations clearly have a number of advantages above and beyond other assessment methods for individuals with PIMD. As a consequence of the severity of their disabilities, individuals with PIMD are not able to use spoken language, and so they express themselves by means of body language. Consequently, individuals in the target group often cannot fulfill the requirements of standardized assessment instruments in terms of motor and speech abilities (Vlaskamp, 2005). The communication of individuals with PIMD mostly consists of subtle signals that are difficult to detect for DSPs (Wilder & Granlund, 2003). The same signal may have a different meaning for different individuals (Vlaskamp). While physiological measurements can help to register these subtle signals, the results do not reveal the necessary information about the meaning of these signals for individuals in the target group (Mudford et al., 1997). Looking at the individual's reactions in different situations, DSPs can learn to interpret the different kinds of behavior (Grove, Bunning, Porter, & Olsson, 1999). Only observations allow DSPs to take the meaning of the individual's behavior into account. Detailed registration of the behavior and, at the same time, of the influencing factors are especially important when observing individuals with PIMD.

However, general problems are related to observations in individuals with PIMD as well. Observations often lack an unambiguous description of their focus. When observations are based on theoretical concepts, these cannot be directly linked to visible behavior. Consequently, observers are forced to interpret the visible behavior, and ascribing meaning to behavior is, in turn, always interpretation (Vlaskamp, 2005). Interpretation, then, can be specified as yet another problem to do with observations. While it is important for DSPs to take the meaning of the behavior of their clients into account, several factors can bias the interpretation. DSPs interpret the behavior of the individual with PIMD based on their knowledge of the individual and previous experiences with the individual in similar situations. Since this knowledge differs for each DSP, observations of the same situation may result in different scores (Grove et al., 1999).

Additionally, general expectations of reactions and contextual factors in a specific situation can also influence the DSP's judgment (Hogg, Reeves, Roberts, & Mudford, 2001). While DSPs' overall judgments remain similar in situations with and without contextual information, DSPs judge the individual's expressions more positively when they expect the individual with PIMD to enjoy an activity than when they do not know about the content of the activity. Furthermore, and as a consequence of the subjectivity of the interpretations, observations of individuals with PIMD regularly result in low reliability. As a result, researchers find themselves still involved in discussions about influencing factors and explanations (Vlaskamp).

The general problems that we experience in observations of individuals with PIMD also become apparent in the alertness observations of individuals in the target group. In the literature, no unambiguous description of alertness has been found (Munde et al., 2009). Although the authors all agreed that it was possible to observe alertness in the behavior of individuals with PIMD, different terms with different descriptions were found to have been introduced. Additionally, different scoring categories were used to determine alertness levels. Another point of discussion is scoring frequency. Because of quick and irregular changes in alertness levels, some authors plead the case for continuous scoring (Guess et al., 1999; Mudford et al., 1997). However, the difference in content information based on interval scoring is not yet made evident here, and, above all else, it should be remembered that interval scoring is actually more useful in clinical practice. To measure and compare the impact of these differences, researchers found themselves obliged to determine the reliability of their observations. For a number of the studies, reliability did not exceed the formulated criterion (Mudford et al.; Woodyatt, Marinac, Darnell, Sigafos, & Halle, 2004). Although different explanations for these results have been discussed (Arthur, 2000; Guess, Roberts, Behrens, & Rues, 1998; Mudford, Hogg, & Roberts, 1999), no solution for the problem of low reliability in alertness observations has been offered.

Taking the above-mentioned problems into account, the Alertness Observation List (AOL)

has been developed accordingly (Vlaskamp, Fontaine, & Tadema, 2005). Within the AOL, a clear description of alertness is employed. Alertness is described as the “level” of being open to the environment. DSPs are thus able to use the AOL to formulate an individual alertness profile. Alertness is scored on four different “levels” in order to search for alertness patterns over the period of a day and to find out about changes in alertness based on the impact of different stimuli. Thereby, scoring frequency increases for each of the three subsequent scoring forms of the AOL.

The aim of the present study was to estimate the reliability of the AOL. In a previous study, the AOL was proved to be reliable in five cases (Petitiaux, Elsinga, Cuppen-Fontaine, & Vlaskamp, 2006). In the present study, we determined the general reliability of the instrument for a larger sample. In doing so, we strived to reach adequate reliability results while, at the same time, minimizing the problems with observations mentioned above.

Method

Participants and Setting

A Dutch school for special education volunteered to use the instrument. In this school, four classes were randomly selected. All 23 students of the four classes (12 girls and 11 boys) were included in the study. The children’s ages ranged from 6 to 16 years ($M = 11, 57, SD = 3, 25$). All the children could be described as individuals with PIMD. In addition to profound intellectual and profound motor disabilities (Nakken & Vlaskamp, 2007), individuals in the target group suffer from additional sensory impairments and a broad range of health problems (e.g., epilepsy, dysphagia, constipation, gastro-oesophageal reflux, Arvio & Sillanpää, 2003; Kapell et al., 1998; Van Schrojenstein Lantman-de Valk, van den Akker, Maaskant, & Haveman, 1997; Van Splunder, Stilma, Bernsen, & Evenhuis, 2006). For the children involved in the present study, the diagnoses included a number of different syndromes such as West’s syndrome and Battered Child Syndrome. For 60% of the children, no clear medical diagnosis had been formulated. Visual and auditory impairments were assessed in 57% and 13%

of the children, respectively. Additionally, 22% of the participants suffered from epilepsy. For all the children, informed consent to take part in this study was obtained from their parents or legal representative.

To take into account the possible impact the observers’ knowledge might have and, as a result, to determine reliability as objectively as possible, the observations were conducted by three types of observers: teachers, an external observer who had received additional information about the children and an external observer who did not know the children with PIMD at all. All the observers were familiar with the AOL and were aware of the aim of the research.

Instrument

The Alertness Observation List was used to determine alertness. In the AOL, four “levels” of alertness are distinguished: 1) active, focused on the environment; 2) inactive, withdrawn; 3) sleeping, drowsy; and 4) agitated, discontented. Each “level” is assigned a color: green, orange, red and blue, respectively. More detailed descriptions of the different “levels” are given in Table 1. Four different forms are used to develop a complete alertness profile. The first form was completed before starting the observations. The overall state of the individual on the day before the planned observation and on the day of the observation itself was determined. If the individual had recently been ill or had had an unusual epileptic seizure the same day, observations were not conducted. The second form was used to observe an individual for three days, scoring alertness every 15 minutes, starting when the individual entered the school and stopping when he or she left. Before using the third form, DSPs first chose the optimal moment in a day for offering an educational activity to the individual. The stimulation was then presented for 15 minutes with the precondition that the score had to have been green or orange, thus that the individual had been awake during the preceding 15 minutes. The alertness “level” was scored every five minutes during the activity and for 15 minutes following that (Form 3). The fourth form was similar to the third one, except that the observer scored every 20 sec-

TABLE 1

Descriptions of the Different Alertness “levels” According to the AOL (Vlaskamp, Fonteine, & Tadema, 2005)

<i>Alertness “level”</i>	<i>Color</i>	<i>Description</i>	<i>Examples of behavior</i>
Active, focused on the environment	Green	The individual is engaged in sensory activities. That means he or she is looking, listening, touching or smelling. These activities are directed toward the environment. As a result, the individual is able to be focused on other individuals or on materials in the room.	Open eyes, focusing with the eyes, turning the head or the eyes in the direction of a stimulus, the body is tensed, the individual is reaching or grasping an object, the individual is eating or drinking.
Inactive, withdrawn	Orange	The individual may be engaged in sensory or motor activities. These activities are <i>not</i> directed to the environment. Activities can be focused on the individual him/herself or without any focus.	Looking at one’s hands, stereotypical movements, sensing one’s clothes, staring, fiddling with one’s body, the head down or turned aside, thumb-sucking, groaning softly, rubbing the eyes, rolling the head, rocking him/herself.
Sleeping, drowsy	Red	The individual is sleeping. Movements and sounds correspond to sleep.	Sleeping, eyelids are shut, eyes are opening and closing slowly, snoring, limbs are limp and loose.
Agitated, discontented	Blue	The individual expresses discomfort.	Shouting, crying, screaming, hitting or kicking materials or persons, banging with the head, hitting, biting, scratching, or kicking himself or herself.

onds for a period of five minutes, during which the child was offered the activity. Finally, using all the information gathered on the observation forms, an alertness profile could be formulated and written down in a “traffic light.” This overall description of all the alertness categories of the individual was complemented with concrete examples of behavior for each category (Vlaskamp et al., 2005).

Procedure

The AOL was completed for all 23 children. Since the first three forms of the AOL are conditional relative to Form 4, the fourth form is the one that is expected to reveal the most relevant information. Consequently, we have only included the observations of the fourth form in the present study. All observations using this form were videotaped. The

observations were conducted in the classroom and in a multisensory room. In these two different settings, five and two observations, respectively, were completed for each child. A number of videotapes were excluded from the study because of the low quality of the recordings. The remaining pool of observations consisted of 120 situations.

To investigate the reliability of the AOL, 39 situations were randomly selected from the pool of 120 observations. For every situation, we asked two observers to score alertness for the individual with PIMD. Using the general formula for agreement, we estimated the inter-observer agreement. Since we investigated an individual judgment, a minimum value of 80% was applied in order to interpret the results (Mudford et al., 1997).

Additionally, the intra-observer agreement was calculated for another 39 situations. The situations were again selected at random from

the pool of 120 observations. Observers were asked to complete the observation form twice for the same situation, six weeks apart. The same formula and the same interpretation criteria as for the inter-observer agreement were used to estimate and interpret agreement percentages.

Results

All 78 situations were scored using the fourth form of the AOL. Employing the general formula for agreement in order to calculate the inter-observer reliability for 39 of the situations, we found $r = 81\%$ ($Mdn = 81.44$; $M = 81.46$; $SD = 13.88$). The intra-observer reliability for the other half of the situations was $r = 87\%$ ($Mdn = 86.79$; $M = 85.23$; $SD = 11.75$). Although the median exceeded the formulated criterion of 80%, individual results showed large differences. Inter-observer reliability ranged from 50% to 100% with a standard deviation of 13.88. These results were similar to those for the intra-observer reliability that had a range from 61.11% to 100% and a standard deviation of 11.75. In addition, those observers who received more information about the children scored higher results for the inter-observer reliability, whereas the scores of the observer who did not know the children at all were higher for the intra-observer reliability. An overview of all the results including the percentages for each situation and each type of observer is presented in Tables 2 and 3.

Conclusion and Discussion

The present study does show that the AOL is a reliable instrument for determining alertness in individuals with PIMD. However, a number of details need to be discussed. Although the overall results of the present study are sufficient, the large range of results for the different situations is striking. The differences can be partially explained by the severity of the disabilities of the target group. One example of this is that visual impairments are common in people with PIMD, and so those in the target group with visual impairments most likely will not show their focus on the environment by directing their eyes or head. It is therefore difficult to determine their alertness

levels. Another example is that uncontrolled movements of individuals with spasticity also aggravate alertness observations. When individuals are not able to show their focus by pointing or grasping, DSPs might well interpret the individual's behavior as being "not alert." The frequency of changes in alertness "levels" is another explanation for the large individual differences in reliability. When the individual with PIMD showed a clear focus during the entire observation, reliability was always 100%. In contrast, frequent changes in alertness "levels" were associated with lower reliability. As Guess et al. (1999) and Mudford et al. (1997) found in previous studies, the differentiation between orange and green alertness "levels" was especially difficult. Therefore, observations concerning situations with numerous changes between these two "levels" can lead to low reliability results.

Individual differences in terms of alertness expressions may also have an impact on the reliability of the results. Although observers are expected to take these differences into account, idiosyncratic behavior can aggravate observations of individuals in the target group (Hogg et al., 2001). Turning away the head may be an indication of dislike for the stimulus presented, but, by the same token, an individual with visual impairments may also express his or her interest in this way, especially as a reaction to an auditory stimulus. In such situations, the proxies' knowledge of the children may be seen as an advantage in interpretation of their behavior. However, looking at the higher intra-observer reliability for the observer who did not know the children at all, we are obliged to amend this statement. Since external observers were not influenced by their knowledge about the child and recent experiences with the child, their observations were mainly based on the observable behavior and their interpretations were actually more consistent. However, there is no real standard for judging the correctness of the interpretations. Consequently, observations by proxies and external observers might well be used to greater advantage as complementary sources of information (Petry & Maes, 2006).

The present study has confirmed that observing individuals with PIMD reliably is an enduring challenge for DSPs and researchers. The subjectivity of the interpretations remains

TABLE 2
Results for Inter-Observer Reliability

<i>Observers</i>	<i>Situations</i>													<i>M</i>	<i>Mdn</i>	<i>SD</i>	
	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>	<i>10</i>	<i>11</i>	<i>12</i>	<i>13</i>				
T and E+	83.33	75.00	87.50	75.17	80.00	100.00	75.00	62.50	100.00	87.50	62.50	100.00	87.50	87.50	82.77	83.33	12.80
T and E-	50.00	75.00	70.83	79.17	83.33	91.67	100.00	100.00	87.50	91.67	79.16	66.67	75.00	75.00	80.77	79.17	13.98
E+ and E-	83.33	80.00	86.67	75.86	93.33	100.00	50.00	64.86	81.82	94.74	64.44	100.00	76.00	76.00	80.85	81.82	14.86
													TOTAL	TOTAL	81.46	81.44	13.88

Note: The letters refer to the different types of observers. Teachers are designated by a "T," the external observer who received additional information about the children is designated by an "E+" and the external observer who did not know the children at all is designated by an "E-." The situations were not similar for the different observers.

TABLE 3
Results for Intra-Observer Reliability

<i>Observers</i>	<i>Situations</i>													<i>M</i>	<i>Mdn</i>	<i>SD</i>	
	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>	<i>10</i>	<i>11</i>	<i>12</i>	<i>13</i>				
T	70.83	70.83	91.67	100.00	79.17	83.33	70.83	70.83	100.00	70.83	100.00	91.67	83.33	83.33	83.33	83.33	12.15
E+	95.83	66.67	76.67	93.10	93.55	100.00	88.89	61.11	87.10	78.95	86.36	100.00	85.00	85.00	85.63	87.10	12.02
E-	91.07	96.77	63.64	72.88	95.16	82.98	100.00	90.48	91.53	86.05	97.73	86.84	72.22	72.22	86.72	90.48	11.09
													TOTAL	TOTAL	85.23	86.97	11.75

Note: The letters refer to the different types of observers. Teachers are designated by a "T," the external observer who received additional information about the children is designated by an "E+" and the external observer who did not know the children at all is designated by an "E-." The situations were not similar for the different observers.

a threat to their reliability. Additionally, several practical problems aggravate the implementation of alertness observations for individuals in the target group. The scoring systems are experienced as being too complex for use in clinical practice. The observations themselves are found to be time-consuming as well and, therefore, often not practicable in day-to-day situations (Petry et al., 2006). However, based on the example of the AOL, we can conclude that reliable observations are in fact possible. While alertness observations remain an effort, DSPs also confirmed the importance of determining alertness in their clients. Furthermore, investigating the value of the results of the “traffic light” for DSPs may reveal additional information about the usefulness of the AOL in clinical practice.

References

- Arthur, M. (2000). Behavior states and a half-full glass: A response to Mudford, Hogg, and Roberts. *American Journal on Mental Retardation, 105*, 509–511.
- Arvio, M., & Sillanpää, M. (2003). Prevalence, etiology and comorbidity of severe and profound intellectual disability in Finland. *Journal of Intellectual Disability Research, 47*, 108–112.
- Grove, N., Bunning, K., Porter, J., & Olsson, C. (1999). See what I mean: Interpreting the meaning of communication by people with severe and profound intellectual disabilities. *Journal of Applied Research in Intellectual Disabilities, 12*, 190–203.
- Guess, D., Roberts, S., & Guy, B. (1999). Implications of behavior state for the assessment and education of students with profound disabilities. In A. C. Repp & R. H. Horner (Eds.), *Functional analysis of problem behavior – from effective assessment to effective support* (pp. 338–394). Belmont: Wadsworth.
- Guess, D., Roberts, S., Behrens, G. A., & Rues, J. (1998). 'Interobserver agreement and disagreement in continuous recording exemplified by measurement of behavior state': Comment. *American Journal on Mental Retardation, 103*, 75–79.
- Guess, D., Roberts, S., Siegel-Causey, E., Ault, M., Guy, B., & Thompson, B. (1993). Analysis of behavior state conditions and associated environmental variables among students with profound handicaps. *American Journal on Mental Retardation, 97*, 634–653.
- Hogg, J., Reeves, D., Roberts, J., & Mudford, O. C. (2001). Consistency, context and confidence in judgements of affective communication in adults with profound intellectual and multiple disabilities. *Journal of Intellectual Disability Research, 45*, 18–29.
- Kapell, D., Nightingale, B., Rodriguez, A., Lee, J. H., Zigman, W. B., & Schupf, N. (1998). Prevalence of chronic medical conditions in adults with mental retardation: Comparison with the general population. *Mental Retardation, 36*, 269–279.
- Mudford, O. C., Hogg, J., & Roberts, J. (1997). Interobserver agreement and disagreement in continuous recording exemplified by measurement of behavior state. *American Journal on Mental Retardation, 102*, 54–66.
- Mudford, O. C., Hogg, J., & Roberts, J. (1999). Behavior States: Now You See Them, Now You Don't. *American Journal on Mental Retardation, 104*, 385–391.
- Munde, V. S., Vlaskamp, C., Ruijsenaars, A. J. J. M., & Nakken, H. (2009). Alertness in individuals with profound intellectual and multiple disabilities: a literature review. *Research in Developmental Disabilities, 30*, 462–480.
- Nakken, H., & Vlaskamp, C. (2007). A need for a taxonomy for profound intellectual and multiple disabilities. *Journal of Policy and Practice in Intellectual Disabilities, 4*, 83–87.
- Nelson, C., van Dijk, J., McDonnell, A. P., & Thompson, K. (2002). A framework for understanding young children with severe multiple disabilities: The van Dijk approach to assessment. *Research and Practice for Persons with Severe Disabilities, 27*, 97–111.
- Petitiaux, W. S. D., Elsinga, G. M., Cuppen-Fontaine, H., & Vlaskamp, C. (2006). Alertheid van kinderen met ernstige verstandelijke en meervoudige beperkingen: een casestudy. [Alertness in children with profound intellectual and multiple disabilities: a case study]. *NTZ, 32*, 241–254.
- Petry, K., & Maes, B. (2006). Identifying expressions of pleasure and displeasure by persons with profound and multiple disabilities. *Journal of Intellectual & Developmental Disability, 31*, 28–38.
- Van Schrojenstein Lantman-de Valk, van den Akker, M., Maaskant, M. A., & Haveman, M. J. (1997). Prevalence and incidence of health problems in people with intellectual disability. *Journal of Intellectual Disability Research, 41*, 42–51.
- Van Splunder, J. A. N. G., Stijlma, J. S., Bernsen, R. M. D., & Evenhuis, H. M. (2006). Prevalence of visual impairment in adults with intellectual disabilities in the Netherlands: cross-sectional study. *Eye, 20*, 1004–1010.
- Vlaskamp, C. (2005). Assessing people with profound intellectual and multiple disabilities. In J. Hogg & A. Langa (Eds.), *Assessing adults with intellectual disability: A service provider's guide* (pp. 152–163). Oxford: Blackwell Publishers.
- Vlaskamp, C., Fontaine, H., & Tadema, A. (2005).

Manual of the list 'Alertness in children with profound intellectual and multiple disabilities'. [Handleiding bij de lijst 'Alertheid van kinderen met zeer ernstige verstandelijke en meervoudige beperkingen']. Groningen: Stichting Kinderstudies.

Wilder, J., & Granlund, M. (2003). Behaviour style and interaction between seven children with multiple disabilities and their caregivers. *Child: Care, Health and Development*, 29, 559–567.

Woodyatt, G., Marinac, J., Darnell, R., Sigafoos, J., & Halle, J. (2004). Behaviour State Analysis in Rett Syndrome: Continuous data reliability measurement. *International Journal of Disability, Development and Education*, 51, 383–400.

Received: 18 August 2009

Initial Acceptance: 16 October 2009

Final Acceptance: 4 December 2009