

Systems-Level Assessment of Interobserver Agreement (IOA) for Implementation of Protective Holding (Therapeutic Restraint) in a Behavioral Healthcare Setting

Ava E. Kleinmann · James K. Luiselli ·
Florence D. DiGennaro · Gary M. Pace ·
Serra R. Langone · Cynthia Cochran

Published online: 21 October 2009
© Springer Science + Business Media, LLC 2009

Abstract Increasing demands for accountability in behavioral services have required data-based and reliable clinical decisions. The importance of reliable data is magnified in settings where restrictive procedures (e.g., therapeutic restraints) and challenging antecedent behaviors (e.g., aggression) are part of the typical treatment day. While reliable data contribute to valid decision making, a number of challenges exist to obtaining these data including, but not limited to, program wide logistics and uncertainty about the independence of observers. The present study acquired interobserver agreement (IOA) data for therapeutic restraints across an entire program day. Data revealed that staff compliance with the recording protocol was high and that a large proportion of therapeutic restraints were captured for IOA purposes. High levels of IOA were also recorded for a number of dimensions of therapeutic restraint implementation. Implications for future research as well as clinical applications are discussed.

Keywords Interobserver agreement · Reliability · Therapeutic restraint

A. E. Kleinmann · J. K. Luiselli · F. D. DiGennaro · G. M. Pace · S. R. Langone · C. Cochran
May Institute, Randolph, MA, USA

A. E. Kleinmann (✉)
Department of Psychology, Western New England College, Sleith Hall #5002, 1215 Wilbraham Road,
Springfield, MA 01119, USA
e-mail: ava@wnec.edu

Many human services and behavioral healthcare organizations rely on data collection to evaluate outcomes (Reid 2004). Today's climate of accountability in providing human services demands objective, data-based decision-making. Objective measurement is critical for documenting progress and making programmatic decisions about the types and intensity of therapeutic procedures. The frequency of criterion-referenced behaviors, percent of achieved learning objectives, and occurrence of sentinel events are some of the dependent measures typically recorded.

For data to be interpretable it must be recorded reliably. Reliability, or interobserver agreement (IOA), refers to the consistency of measurement over time (Hartmann 1977; Kazdin 1977). IOA is assessed by calculating agreement between the simultaneously and independently recorded data sets of two individuals. Although there are several methods for computing IOA (Hartmann 1977), convention stipulates that mean agreement should equal or exceed 85%.

The importance of IOA assessment when collecting and evaluating data is heightened when considering extraordinary, and sometimes controversial, intervention procedures. For example, protective holding (therapeutic restraint) may be required for some children and adults who have serious challenging behaviors such as aggression, self-injury, and property destruction (Harris 1996; Sturmey et al. 2005). Protective holding consists of one or more staff physically immobilizing a person's voluntary movement through restriction of arms and/or legs in a standing, sitting, or supine position. In some cases, protective holding is implemented to manage an unanticipated crisis situation. The procedure also has been employed strategically as one component of a behavior support plan (Luiselli et al. 2000).

Several studies have demonstrated positive effects from protective holding in clinical application with children and adults who have developmental disabilities, mental retardation, brain injury, and psychiatric disorders (Luiselli et al. 2000; Matson and Keyes 1988; Rapoff et al. 1980; Rolider et al. 1991). Despite these findings, protective holding is an invasive procedure that can be misapplied, cause injury (Hill and Spreat 1987; Spreat et al. 1986), is socially stigmatizing, and may not be acceptable to service providers (McDonnell and Sturmey 2000). For these reasons, reliable data recording of protective holding is a necessity in order to document effectiveness and justify continued implementation. Importantly, IOA assessment contributes to intervention integrity by verifying that staff consistently adhere to behavior support plan guidelines (Coddling et al. 2005). In addition to improving clinical service delivery, IOA assessment of protective holding is valuable as an approach toward risk management (Luiselli and Russo 2005).

The previous concerns notwithstanding, performing routine IOA assessment can be an obstacle for large-population service organizations. In fact, we were unable to locate any published descriptions of systems-wide all day approaches to recording IOA. With regard to protective holding, studies have addressed antecedent evaluation (Luiselli et al. 2003), hold reduction procedures (Luiselli 2009; Luiselli et al. 2005), and social validity (Cunningham et al. 2003; McDonnell and Sturmey 2000), but not IOA assessment. Accordingly, the purpose of the present research was twofold. First, we designed a method for recording all day IOA by educational staff at a specialized school for children and adolescents with traumatic brain injury (TBI). A second objective was to apply the IOA assessment protocol to protective

holding interventions. Specifically, we evaluated assessment data to determine whether direct-care staff consistently identified the antecedent behaviors leading to protective holding and implemented the procedure as instructed in a student-specific behavior support plan. In summary, the study represents a systems-level application of behavior analysis with clinical implications for service centers where protective holding is an element of therapeutic programming.

Method

Setting

The setting was a private day school for children aged 7–21 with brain injury, intellectual disability, and challenging behaviors. The school subscribed to a behavioral model of service-delivery replete with individualized behavior support plans for all students, program-wide data systems, frequent staff training, and a team of behavior specialists who provided ongoing consultation. The study incorporated nine classrooms that served a total of 46 students, with ratios approximating two students to one staff person. At the onset of the study, 26 (56.5%) students had a protective hold or an escort (i.e. moving hold) in their behavior support plan.

Participants

Classroom and administrative staff, including senior-level teachers (22.4%), assistant teachers (9%), direct care staff (52.1%), and others (e.g., administrators, transitional education specialists; 35.4%) functioned as observers. Participating observer dyads represented three general categories of professional relationships: supervisor to supervisee (e.g., senior-level teacher to direct care staff; 61.5%), peer to peer (e.g., two direct care staff; 29.2%), and others (e.g., teacher to vocational staff; 9.4%). Due to the anonymity of data collection, the number of participating individual staff represented was unable to be determined. Nineteen individual students were represented among the completed forms (described below), with an average of 10.1 observation pairs per student (range = 1–49). Eight of the nine classrooms participated (one did not have any protective holds or escorts during the study), with an average of 24 observation pairs per classroom (range = 1–60). Procedural integrity checks were completed by nine individuals ($M=7.6$ checks, range = 1–27).

Measures

IOA Form [Appendix A]. A protective hold was defined as staff using a program approved procedure to physically immobilize a student's voluntary movement through restriction of arms and/or legs in a standing, sitting, or supine position. An escort was defined as a moving hold used to transport a student to a safe area. A hold/escort IOA form was created to capture staff agreement on the type of procedure implemented (i.e., hold, escort), the circumstance under which the procedure was implemented (i.e., per plan, emergency), and the antecedent behavior

to the procedure (e.g., tantrum, aggression). The same form was used for all students, and respondents wrote in the name, date, and time of the protective hold/escort and then checked off fields addressing the various dimensions of their observation. Respondents also noted their job titles so that the agreement for dyads with different types of professional relationships could be examined. IOA was calculated by the following formula: agreements divided by agreements plus disagreements, multiplied by 100%.

Integrity Form [Appendix B]. A six-step procedural integrity check was developed to examine the extent to which staff completed IOA forms independently. Procedural fidelity was assessed through direct observation of two classroom staff completing the IOA forms. Procedural fidelity was calculated by dividing the number of integrity steps observed by the total number of steps (6), multiplied by 100%.

Procedure

Prior to collection of IOA data, staff were trained on the IOA data collection procedure by the project administrator (the first author) and the administrative team. The data collection procedure included the following: (1) completion of an IOA form when at least two staff observed a protective hold/escort and the antecedent behavior, (2) staff withholding conversation about the form until it was fully completed, with the exception of reminding the other to fill it out or write the same time and date, (3) contacting the administrator on call to report the details of the hold/escort as per existing agency protocol (at that time administrators provided an additional prompt for staff to complete the IOA form), and (4) placement of the IOA forms in an envelope in a pre-specified location.

Administrators were informed of the implementation of a protective hold/escort by the administrator on call, classroom staff, or by directly witnessing its application. Throughout the course of data collection the project administrator collected IOA forms from the envelopes, provided feedback about any possible errors in protocol, and frequently reminded staff to complete the forms. Data were collected over the course of approximately four months (80 school days). The system administrator faded prompts to collect data from almost daily to approximately once per week by the end of data collection.

Forms were excluded based on the following criteria: (1) no obvious match, (2) a discrepancy of ten or more minutes in the times recorded, (3) knowledge that the forms were completed more than one day following the observed protective hold/escort, and (4) partial completion of the form.

Results

A total of 447 forms was returned. One hundred ninety-two pairs were identified based on correspondence of time, date, and student, and were used to calculate IOA. Fifty-nine (13.2%) forms were unable to be matched because of significant discrepancies in the times reported (greater than 10 min apart), and some forms

could not be matched to a second observer. Accordingly, these forms were excluded from the analyses. An additional four forms were included that represented a third observer of the protective hold or escort. The IOA for the four months of data collection represented 47.8% of all protective holds/escorts implemented at the school.

Nearly perfect IOA (99.5%) was found for whether a protective hold or escort occurred, suggesting that staff were able to discriminate between the procedures. There was 98.4% agreement on whether the protective hold or escort occurred per plan, with instances of disagreement coinciding with different observed antecedent behaviors. In 90% of the observation pairs, respondents reported the same antecedent to the protective hold or escort. Further analysis of the disagreement revealed that many of the behaviors endorsed were not mutually exclusive (e.g., noncompliance and tantrum). The high rates of IOA were consistent across the various professional dyads comprising the sample.

Procedural integrity data were collected for 35.4% ($n=68$) of the protective holds and escorts that were captured in the sample. Procedural integrity data revealed high levels of adherence to the IOA data collection procedures (99.3%). In order to minimize the amount of data that had to be excluded, procedures incorporated staff conferring on dates and times. In one instance, the dyad completing the forms did not check with each other on the exact time of the incident which contributed to a lower integrity score. However, this missed step did not undermine the independence of the data which is of primary importance in this study.

Discussion

Collecting IOA data is the standard in behavioral research, where the results from carefully controlled experiments must be evaluated, in part, based on the reliability of measurement. Typically, a minimum of 30% of the sessions are subject to observation by an independent observer. In light of the barriers to reliable recording (see Kazdin 1977), accurate measurement is important to demonstrate. Increasing recognition of the need for documentation of reliability of data has been mirrored by a trend towards including IOA data in journal articles. In some instances manuscripts may not be accepted based on the inadequate levels or absence of reliability data. Extending such standards to clinical settings is a clear application of an important practice in the research arena.

Use of IOA is not merely an academic exercise appropriate for research. Monitoring for reliability of data at a systems level is an important clinical tool in settings where behavioral data are the basis for clinical decision making. The present investigation sought to apply these techniques at a systems level for the most restrictive procedures. Data revealed that the IOA recording protocol was relatively easy to implement, was useful as a clinical tool, and based on informal feedback was satisfactory to staff. Published programs outlining alternative methods of collecting IOA data for an entire program day are non-existent, and the present effort represents a unique contribution to the field. In today's climate of accountability in providing professional services, empirically supported clinical practice is essential.

In the present application, the IOA protocol was implemented across an entire school program and day, with high levels of compliance by staff. Nearly half of all protective holds implemented were captured school-wide. This was particularly significant because many of the protective holds and escorts were not witnessed by more than one staff, rendering these events unattainable for IOA. The high rates of return on the data are also indicative of the practicability of these procedures. The IOA forms were created so that staff requirements were minimal. Respondents completed a brief checklist designed to be finished in under one minute. Given the multiple demands on staff time and in recognition of the aversiveness of paperwork, this was an important design element to incorporate. Sustaining a procedure of this type is promising in light of these factors.

Another notable finding was the ability of a systems wide intervention during an all day program to capture a sizeable proportion of all protective holds. IOA for one-third of all protective holds was consistent with convention regarding the minimum proportion of sessions for which reliability data should be taken. It is expected that as the procedures become more familiar and automatic to staff, increasing rates of return of forms will follow. Also important to consider are the roles of the IOA system administrator and behavior team. Prompting of staff by the team was integral to the high rates of return. On some level, a negative reinforcement contingency may have been operating in that intensity of staff prompting was indirectly related to their compliance with IOA procedures.

It is also important to note that the IOA procedures were layered into existing staff protocol following the implementation of protective holding. Existing procedures required staff to call an administrator following the implementation of a protective hold or escort and to document the event in writing. Both calls and paperwork were logged administratively to monitor program wide implementation of physical intervention. Accordingly, built in checks on paperwork for every call was a component of the system. In addition, when staff reported a protective hold this was a natural opportunity for the supervisor to prompt staff to complete IOA forms. The multiple levels of involvement of various staff in the documentation of protective holds/escorts allowed for increased collaboration and systems checks for the most restrictive procedures implemented in the setting. Also notable was that the IOA protocol was introduced and monitored by existing staff and did not require the hiring of additional personnel.

The levels of IOA indicated by the data also demonstrated that staff independently agreed on a number of aspects of the protective holds/escorts that were reported. Staff agreed when the procedures were implemented, and they were almost always implemented per plan. The finding of high levels of concordance was replicated across all staff dyads (e.g., supervisor to supervisee; peer to peer). The importance of high IOA is particularly important with these procedures given their restrictiveness, inherent risks, and stricter guidelines governing their implementation. At the same time, some discrepancies in behaviors occasioning protective holds/escorts were documented and highlighted the areas where further staff training was warranted.

As such, a number of recommendations for the clinical application of these types of IOA data emerged from the findings. While the applications described here are

specific to our setting, the various levels of data interpretation demonstrate that this simple protocol has several clinical purposes. For example, the high level of IOA across a range of students and classrooms indicated that the program consistently implements restrictive procedures under circumstances where they are warranted and with independent agreement by staff that the same event was witnessed. In fact, on a subsequent review by an accrediting agency, the program received a commendation for the IOA protocol. Maintaining approval to use very restrictive, though sometimes clinically indicated, procedures is critical in a setting such as the one described in this study. Recognizing the importance and utility of this procedure provided social validity for IOA as a systems tool as well.

On other levels the IOA protocol was applied to examine reliability of implementation of protective holds/escorts in certain classrooms. In one instance, the IOA data revealed that for a particular classroom a very small proportion of therapeutic restraints were being captured even though these staff received the same training and level of prompting to complete the forms. Further investigation revealed that many of the protective holds/escorts were implemented by a single staff alone with a consumer. Given that this particular classroom also had a high rate of protective holds it was important to consider staffing patterns both to increase opportunities for IOA and for safety. An additional permanent staff member subsequently was added to this classroom, producing an increase in return rate for IOA forms and modest decrease in protective holds/escorts for some students. Given concerns regarding practicality and ethics of restraint procedures (Sturmey et al. 2005) these data are valuable for ensuring staff and student safety, staff training and follow-up.

The IOA data were also clinically applicable at the level of an individual student. For example, for one student IOA was consistently high for the implementation of protective holds and the circumstances that surrounded their implementation. However, the frequency of challenging behaviors that occasioned them were increasing. Coupled with additional procedural integrity data (as per Coddling et al.'s methodology 2005), the consensus was that the behavior support plan for this student was not producing the desired effect. Subsequent modification of the protective hold criteria followed, with empirical support that was essential to guide the complex clinical decision. In another instance, IOA data for a particular student revealed that staff disagreed on the antecedent behaviors for a significant proportion of the protective holds/escorts that were captured with the IOA procedures. Clarifying and simplifying the operational definitions of these behaviors to staff was therefore warranted to increase staff consistency.

To some extent the data on the implementation of protective holds/escorts served as a crude index for the reliability of data collection on other behaviors that were commonly monitored in the setting (e.g., tantrum, noncompliance). For many consumers protective holds and escorts were part of a complex and escalating behavioral chain, and, as such, the frequencies of these behaviors (and procedures) were somewhat correlated. At the same time, the limitations of using this one index are clear and warrant the collection of additional data. In our own setting the subsequent introduction of IOA procedures for a range of maladaptive behaviors exhibited by consumers has been a useful complement to IOA data for protective holds/escorts.

The present findings must be interpreted in the context of the following limitations. First is the recognition that the study featured a nonrandom sample. Nonreported cases were, in part, comprised of events for which there was no second observer, but may also include events that were more ambiguous and for which staff were hesitant to complete forms. In addition, more experienced staff may have been overrepresented in the sample because of greater familiarity with the protocol and the behaviors that were being reported. The sample data, however, showed that a broad range of students, classrooms, and staff were represented, and that staff varying in terms of their professional relationships and, presumably, their experience agreed at the same rates throughout the course of the study.

Relatedly, it is presumed that staff would prefer to agree with each other when completing paperwork, particularly with regard to restrictive procedures. This may pose a threat to the independence of the data or skew the sample towards events with very clear antecedent behaviors (see Kazdin 1977). While it is impossible to control for every variable across an entire clinical setting, the procedural integrity check incorporated into the IOA protocol provided an index for the independence of the data. In addition, the anonymity of the forms as well as the marketing of the IOA program to staff as a training tool, were design elements incorporated to minimize possible nonrandom responding.

Another limitation was that the same IOA form was used for each student. Using a universal form allowed for easy training and consistency for the basis of aggregating data across a variety of students. However, the antecedent behaviors listed were defined differently for each consumer which may have created some confusion to individuals completing the forms. For example, “aggression” for one consumer may be defined as “unwanted physical contact above the shoulders”, whereas another definition of aggression might include “spitting on any person.” Despite this limitation, agreement on antecedents was still high. This finding provided more support for the identified benefits of using a universal IOA form.

While the findings from this demonstration were positive in terms of the feasibility of a program wide IOA protocol that yielded valuable data, a number of additional questions emerged, particularly with respect to the generalization of the procedures. First, extending these procedures to other behaviors would be helpful to better understand the feasibility of a broader application of the procedures. Similarly, extending the procedures to a residential setting, and particularly one affiliated with a school, would allow for examination of IOA data collection over a 24 h period. Similarly, collecting data on adaptive skills (e.g., functionally-equivalent), replacement behaviors, or academic acquisition would be a natural extension of the present demonstration.

The IOA protocol presented here represents one method for collecting reliability data across a program in a clinically meaningful way. Overall, IOA was high in all areas examined, the recording procedures were fairly easy to implement, and the protocol was satisfactory to staff. The results provide a foundation for further research in this area and the continued use of IOA data not only as a research tool but with clinical applicability as well.

Appendix A

Hold/Escort IOA Sheet

Student _____ **Date** _____ **Time** _____
 (of hold/escort)

What did you observe? ___ Hold ___ Escort

Did you observe the behavior that immediately preceded the hold/escort?

___ Yes ___ No (If no, discontinue – no need to submit the form)

Under what circumstances was the hold/escort implemented?

- ___ Per Behavior Support Plan
- ___ Emergency
- ___ Hold was implemented when it should have been an escort (or vice versa)
- ___ None of the above – this hold/escort should not have been implemented.
- Please explain:

Did a behavior occur that should have led to a hold/escort according to the behavior support plan?

___ Yes ___ No

Which per plan antecedent did you observe?

- ___ aggression
- ___ destruction
- ___ tantrum/outburst
- ___ noncompliance
- ___ other _____
- ___ none

Your position:

- ___ Senior teacher
- ___ Teacher
- ___ DCII
- ___ Other _____

Appendix B

Hold/Escort IOA – Integrity Check

Completed by _____

Date _____

Class _____

Time _____

Student _____

Check each of the following observed for staff pair completing the IOA forms:

Integrity Elements	yes	no	no basis for observation
staff double checked with each other to ensure they had the right date/time	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
staff did not talk while completing forms (aside from asking about exact time/date)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
staff did not sit next to each other while completing forms	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
staff did not compare answers after completing forms	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
staff completed forms before discussing/writing the incident report together	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
staff places forms in the designated envelope	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

References

- Codding, R. S., Feinberg, A. B., Dunn, E. K., & Pace, G. M. (2005). Effects of immediate performance feedback on implementation of behavior support plans. *Journal of Applied Behavior Analysis*, *38*, 205–219.
- Cunningham, J., McDonnell, A., Easton, S., & Sturmey, P. (2003). Social validation data on three methods of physical restraint: views of consumers, staff and students. *Research in Developmental Disabilities*, *24*, 307–316.
- Harris, J. (1996). Physical restraint procedures by managing challenging behaviours presented by mentally retarded adults and children. *Research in Developmental Disabilities*, *17*, 99–134.
- Hartmann, D. P. (1977). Considerations in the choice of interobserver reliability estimates. *Journal of Applied Behavior Analysis*, *10*, 103–116.
- Hill, J., & Spreat, S. (1987). Staff injury rates associated with the implementation of contingent restraint. *Mental Retardation*, *25*, 141–145.
- Kazdin, A. E. (1977). Artifact, bias, and complexity of assessment: the ABCs of reliability. *Journal of Applied Behavior Analysis*, *10*, 141–150.

- Luiselli, J. K. (2009). Physical restraint of people with intellectual disability: a review of implementation reduction and elimination procedures. *Journal of Applied Research in Intellectual Disability*, 22, 126–134.
- Luiselli, J. K., & Russo, D. (2005). Clinical peer review: a description of a comprehensive model in behavioral healthcare. *Behavior Modification*, 29, 470–487.
- Luiselli, J. K., Kane, A., Treml, T., & Young, N. (2000). Behavioral intervention to reduce physical restraint of adolescents with developmental disabilities. *Behavioral Interventions*, 15, 317–330.
- Luiselli, J. K., Pace, G. M., & Dunn, E. K. (2003). Antecedent analysis of therapeutic restraint in children and adolescents with acquired brain injury: a descriptive study of four cases. *Brain Injury*, 17, 255–264.
- Luiselli, J. K., Dunn, E. K., & Pace, G. M. (2005). Antecedent assessment and intervention to reduce physical restraint (protective holding) of children and adolescents with acquired brain injury. *Behavioral Interventions*, 20, 51–65.
- Matson, J. L., & Keyes, J. B. (1988). Contingent reinforcement and contingent restraint to treat severe aggression and self-injury in mentally retarded and autistic adults. *Journal of the Multihandicapped Person*, 1, 141–153.
- McDonnell, A. A., & Sturmey, P. (2000). The social validation of three physical restraint procedures: a comparison of young people and professional groups. *Research in Developmental Disabilities*, 21, 85–92.
- Rapoff, M. A., Altman, K., & Christopherson, E. R. (1980). Elimination of a blind child's self-hitting by response contingent brief restraint. *Education and Treatment of Children*, 3, 231–236.
- Reid, D. H. (2004). Training and supervising direct support personnel to carry out behavioral procedures. In J. L. Matson, R. B. Laud & M. L. Matson (Eds.), *Behavior modification for persons with developmental disabilities: treatments and supports* (pp. 73–99). Kingston: NADD.
- Rolider, A., Williams, L., Cummings, A., & Van Houten, R. (1991). The use of a brief movement restriction procedure to eliminate severe inappropriate behavior. *Journal of Behavior Therapy and Experimental Psychiatry*, 22, 23–30.
- Spreat, S., Lipinski, D. P., Hill, J., & Halpin, M. (1986). Safety indices associated with the use of contingent restraint procedures. *Applied Research in Mental Retardation*, 7, 475–481.
- Sturmey, P., Lott, J. D., Laud, R., & Matson, J. L. (2005). Correlates of restraint use in an institutional population: a replication. *Journal of Intellectual Disability Research*, 49, 501–506.