This is Exhibit " C " referred to in the
affidavit of Cheryl Regehr
sworn before me this
30th day of March 2012

Confidence and Professional Judgment in Assessing Children's Risk of Abuse

Cheryl Regehr, Marion Bogo, Aron Shlonsky and Vicki LeBlanc

Research on Social Work Practice 2010 20: 621 originally published online 11 June 2010
DOI: 10.1177/1049731510368050

Research on Social Work Practice
20(6) 621-628

DOI: 10.1177/1049731510368050
http://rswp.sagepub.com

Confidence and Professional Judgment in Assessing Children's Risk of Abuse

Cheryl Regehr[1], Marion Bogo[1], Aron Shlonsky[1], and Vicki LeBlanc[2,3]

Abstract

Objective: Child welfare agencies have moved toward standardized risk assessment measures to improve the reliability with which child's risk of abuse is predicted. Nevertheless, these tools require a degree of subjective judgment. Research to date has not substantially investigated the influence of specific context and worker characteristics on professional judgment in the use of risk assessment measures. **Method**: This research utilized standardized patients performing in scenarios to depict typical child welfare cases. Ninety-six workers interviewed two "families," completed risk assessment measures, and then participated in interviews regarding their subjective views of their decision making and performance. **Results**: There was considerable variability in risk appraisals. Confidence in risk assessment performance was related to age, acute level of stress, and the worker's perceived ability to engage family members. Confidence in risk assessment was further related to case variables. Confidence was not related to level of risk assessed. **Conclusion**: The variation in risk assessment appraisals in this study, despite at times high rates of worker confidence in their appraisals, speaks to the need for ongoing consultation and increased decision support strategies even among highly skilled and trained workers.

Keywords

child welfare, child abuse, risk assessment, professional judgment, actuarial tools, standardized measures

The objective of risk assessment in child welfare is to identify, from cases referred to child welfare authorities, the subgroup of children at high risk for future abuse or neglect so that action may be taken to prevent it (Baird, Wagner, Healy, & Johnson, 2000; Fluke et al., 1995; Pecora, 1991) and to determine the relative risk for subsequent maltreatment in children who have been abused (English & Pecora, 1994; Miller, Williams, English, & Olmstead, 1987). Risk assessment instruments provide a mechanism for targeting treatment resources to the most high-risk children through the use of explicit criteria and uniform approaches to assessment (DePanfilis & Zuravin, 2001; English & Pecora, 1994; Johnson & L'Esperance, 1984). The importance assigned to risk assessment in child welfare practice has sparked considerable controversy, however. Many standardized risk assessment tools and protocols have been implemented in child welfare contexts throughout the world despite the fact that they have questionable reliability and/or validity (Lyons, Doueck, & Wodarski, 1996). Postimplementation studies have yielded mixed results regarding their predictive ability (Baird et al., 2000; Camasso & Jagannathan, 2000; Cicchinelli, 1991; English & Pecora, 1994; McDonald & Marks, 1991; Pecora, 1991; Wald & Woolverton, 1990). Efforts to increase the reliability of these instruments using a greater number of objective items have yielded some success (Baird, et al., 2000) but the final instruments, by the very nature of the phenomenon being predicted, contain elements and processes that require a certain level of subjective judgment (Gambrill & Shlonsky, 2000).

This concept of subjective judgment in the application of actuarial tools is central to our understanding of evidence-based assessment methods. That is, evidence-based practice is the use of best available evidence in concert with client state/circumstances and preferences/values (Gibbs, 2003; Sackett et al., 2000). Clinical expertise can be seen as the optimal integration of these three ingredients of practice (Gambrill, 2006; Haynes, Devereaux, & Guyatt, 2002; McCracken & Marsh, 2008; Meehl, 1954). In child protection, practitioners

[1] Factor-Inwentash Faculty of Social Work, University of Toronto, Ontario, Canada
[2] Faculty of Medicine and Factor-Inwentash Faculty of Social Work, University of Toronto, Ontario, Canada
[3] The Wilson Centre, University of Toronto, Ontario, Canada

Corresponding Author:
Cheryl Regehr, Office of the Provost, University of Toronto, 27 Kings College Circle, Toronto, M5A 1S1, Ontario, Canada
Email: Cheryl.regehr@utoronto.ca

can use reliable and valid tools to assist them in the difficult choices among often competing goals such as keeping the family together versus taking the child into care. This decision-making process requires the application of clinical judgment to standardized risk assessment measures (Shlonsky & Wagner, 2005; Schwalbe, 2008). As noted by Shlonsky and Wagner (2005), actuarial risk assessment does not assist in developing case-specific interventions nor does it engage the family in cooperative case planning. These aspects of risk assessment require clinical inference or "perceptual measurement" (Van de Luitgaarden, 2009). That is, measurement on actuarial tools is based on the social worker's perception of the client's perception of their situation. This process can be seen to reflect a higher order metacompetency where knowledge, skills, and judgment come together for analysis and decision making in specific cases (Bogo et al., 2006; Kane, 1992).

The academy has a long history of examining the concept of professional judgment. Goldberg and Werts (1966), for instance, observed that studies of the inferences made by experienced psychologists based on personality test data, revealed little validity for their conclusions. They concluded that the judgments of any single clinician bear no systematic relationship to those of other clinicians even when judging the same patient on the same trait. In child welfare, Lindsey (1992) provides a historical overview of models for evaluating foster care placement decisions for children in child welfare. His review demonstrates that even when there is agreement between workers on diagnosis, the future course for the child may vary substantially depending on which worker makes the decision. He concludes that decisions do not have a sufficient scientific or clinical basis to suggest that placement is a decision that child welfare workers should be called upon to make. More recently, Arad-Davidzon and Benbenishty (2008) found variability in decisions regarding removal of a child among child protection workers provided with the same information in case vignettes. Thus, even when based on a common understanding of the client's presenting issues (through standardized psychological testing, Diagnostic and Statistical Manual of Mental Disorders (DSM)-driven diagnosis, or standardized case vignettes), clinical judgment is variable.

A number of factors have been proposed to explain differences in clinical judgment (or errors in clinical judgment). For instance, Gambrill (2005) asserts that clinicians can develop a situational awareness and integrate corrective feedback on an ongoing basis, but this requires a degree of self-awareness and critical thinking skills that may not be present despite years of experience. Ambiguous, missing, or contradictory information can also influence judgment (Van de Luitgaarden, 2009), even among highly skilled professionals. Horwath (2006), in a study of social workers in England required to make decisions about child neglect, suggested a number of individual factors that may contribute to disparate decisions regarding risk. These factors included fear of verbal and physical aggression by the parent; fear of making the wrong decision; concerns about lack of skill to identify child neglect; guilt in breaching client trust; and sympathy for families in situations of hardship. As a result, of these and other similar findings, some suggest that professional judgment is more of a subjective moral stance than an objective scientific decision (Chu & Tsui, 2008; Laughlin, 2008).

A different question exists with respect to the certainty or confidence with which a clinician makes a judgment about situations involving risk to a child. Baumann, Deber, and Thompson (1991) identify the possible co-occurrence of micro-certainty with macro-uncertainty, a situation in which a clinician feels certain or confident in a field where no absolute answers exist. Research has demonstrated that individuals working under conditions of uncertainty tend to overestimate the probability (or be overly confident) that their judgments are correct (Smith & Dumont, 2002). Other research cautions that overconfidence can cause people to underestimate the magnitude of their uncertainty and the cost of their error (Mamassian, 2008). Similar to errors in judgment based on heuristic processes (Gambrill, 2006), sources of overconfidence in clinical judgment include (a) confirmatory bias, where an individual attends to information that confirms early impressions; (b) dispositionism, in which people overestimate the impact of personality characteristics and underestimate the importance of situational variables; (c) representativeness, where it is believed that the client represents the classic case; and (d) calibration errors, where the probability of an outcome is underestimated (Arkes, 1981; Baumann et al., 1991; Glascoe & Dworkin, 1993). Not surprisingly, research suggests that confidence in professional decision making increases with experience (Hay et al., 2008); yet, it is not clear whether this is related to the accuracy of judgments.

Although the concept of professional judgment or clinical expertise has been applied to actuarial assessment tools, research to date has not adequately investigated the degree to which specific context and worker variables may influence professional judgment and the manner in which a worker assesses risk even on a "standardized" instrument (Gambrill & Shlonsky, 2000). This research seeks to examine worker confidence when using actuarial measures and factors that influence worker judgment in the use of tools.

Method

This research utilized actors trained as standardized patients, performing in scenarios constructed to depict typical child welfare cases in a 2×2 factorial design. The research protocol was approved by the Research with Human Subjects Ethics Board of the University of Toronto and the project was funded by a national research funding body.

Participants were 96 child welfare workers employed at 12 different child welfare offices located in a large urban center, smaller cities and rural communities. The age range of participants was 22–63 years with a mean age of 35.19 ($SD = 9.1$). Eighty-one percentage of participants were women. Most (62%) were married or lived common law. Forty-eight percentage of participants were intake workers; 34.4% were family service workers; 13.5% held other social work positions; and 4.3% were managers or supervisors. In terms of level of

education, 2.1% had community college diplomas in social service work, 53.1% had BSW degrees, 29.2% had MSW degrees, and the remaining workers had other university degrees.

Research Protocol

Baseline Tests

Participants completed a series of questionnaires aimed at understanding their previous work history and current emotional state.

State-Trait Anxiety Inventory (STAI). State anxiety is a commonly used assessment for stress manipulations, as it has been shown to be sensitive to acute stress manipulations (Spielberger, 1983). The participants completed the state form of the STAI. The State Anxiety (S-anxiety) scale consists of 20 statements (e.g., "I am tense"), to which respondents indicate their level of agreement on a 4-point scale regarding how they feel *at the given moment* (1 = *not at all*; 2 = *somewhat*; 3 = *moderately so*; 4 = *very much so*). The internal consistency of the STAI S-anxiety scale is high, with an α of .92 (Spielberger, 1983). A mean STAI anxiety score of 35 is considered normative for adults of different age groups (Spielberger, 1983). Previous research defines the high-anxiety state at STAI > 45, one standard deviation above the norm (Millar et al., 1995; Moerman, 1996). This measure was administered just after each simulation.

Simulated Child Welfare Risk Assessment

Two 15-min scenarios designed to simulate acutely stressful clinical encounters were developed through a consultative process with experts in child welfare. The scenarios were intended to represent two typical families that a child welfare worker might encounter in their daily work. Once they were prepared, a focus group of 10 child welfare workers were asked to read the scenarios and complete risk assessment forms. They were then asked to comment on whether there was adequate data to complete the standardized risk assessment forms and whether the scenarios represented realistic client encounters. Feedback from this process was used to improve the original scenarios.

The first scenario involved an interview with a mother (Ms. Smith) of an infant following a report by the child's day care provider that welts had been observed on the child. The second scenario involved an interview with the mother (Ms. Samuels) of a latency aged child following the report by a school that the child had disclosed physical abuse. Both scenarios involved White mothers in order to limit the number of variables (such as race and gender) that may influence perceptions of risk. Each scenario was presented in one of two forms, with the parent being confrontational and with the parent being nonconfrontational. Thus, each worker conducted two risk assessment interviews, either with Ms. Samuels as a confrontational client and Ms. Smith as a cooperative client or vice versa. Order of interviews was modified and randomly assigned in a 2×2 design to allow for examination of various order effects.

Standardized patients (SPs) were utilized to portray the role of parents. SPs are healthy individuals, in this case professional actors, trained to portray the personal history, physical symptoms, emotional characteristics, and everyday concerns of actual patients. They are also trained to give consistent information and maintain a consistent level of emotional intensity and engagement. However, the nature and quantity of information they provide is dependent on the questions posed by the interviewer. No deception was used in this study in that participants were aware that the "parents" they were interviewing were actors. Nevertheless, many workers commented on the practice-like experience of the interviews and the ability of the actors to accurately portray parents being questioned about safety issues related to their child.

Completion of Risk Assessment Measures

At the end of each scenario, participants completed three risk assessment measures. These measures include both consensus-based and actuarial-based tools.

The *Ontario Risk Assessment Measure (ORAM)* is a consensus-based child welfare risk assessment instrument used in Ontario from 1998 to 2007. It purports to appraise the likelihood of future harm, expressed as maltreatment recurrence over time but is also inclusive of other perceived harm (e.g., developmental) and aids the worker to assign an overall risk level for maltreatment recurrence. Most accurately described as a clinical judgment, the overall risk rating ranges from 0 (*No/low risk*) to 4 (*High risk*). Ratings made with this measure in laboratory settings tend to be unreliable and the tool does not produce accurate predictions of maltreatment recurrence in actual practice settings (Barber et al., 2008), indicating that this tool should have a substantial degree of variability in ratings. Creating the opportunity for variability in risk score based on the interaction between child welfare workers and SPs was crucial for testing whether stressful interactions would translate into elevated risk scores.

The Ontario Safety Assessment (OSA) is a consensus-based tool that determines whether the child is in immediate danger and consists of questions addressing the caregiver's current and previous behavior, ability to supervise, and attitude toward the child. There are 14 safety threats including items such as child vulnerability and presence of domestic violence in the home for which the child protection worker must indicate whether the threat exists. On the basis of this measure, workers make a final judgment referred to as the safety decision as to whether the child is safe or unsafe.

The Ontario Family Risk Assessment (OFRA) is an actuarial-based instrument that assesses the future risk of maltreatment. Based on the California Family Risk Assessment (Wagner & Johnson, 2006), items are rated and points awarded for the occurrence of specific features such as parental mental health problems and age of the child. A total score is generated for neglect (maximum 16) and abuse (maximum 18), and there

are cutoff scores for a 4-point relative risk scale ranging from *low* to *very high*. This tool has been retrospectively and prospectively validated and appears to successfully differentiate between levels of risk for re-report for maltreatment, substantiated/indicated/verified recurrence of maltreatment, and subsequent entry to foster care (Wagner & Johnson, 2006).

Interviews on Confidence and Decision Making

Following each scenario and completion of risk assessment measures, participants were interviewed to determine their subjective views of their decision making and performance. All 96 participants engaged in two interviews (one following each scenario) resulting in a total of 192 interviews. Each participant was asked to rate their confidence in their performance on a scale of 1–5 and their confidence in their final determination of risk on a scale of 1–5 for each interview. In the interview, they were asked to discuss how they felt about their performance, what they may have done differently, and what information they used in making the assessment of risk. Interview data were transcribed and then divided according to the distributions of scores. Groupings of high confidence, moderate confidence, and low confidence were formed and subjected to thematic analysis. High confidence was defined as Level 5 on the 5-point scale. Because very few participants rated their confidence as Level 1, low confidence was defined as Level 1 and Level 2. This resulted in groupings of high and low confidence for both performance and risk assessment that contained between 12 and 21 interviews for analysis. Only high- and low-confidence groupings were analyzed in order that the contrast between the two could be examined.

Results

Findings of Risk

The OSA asks workers investigating child maltreatment to make a judgment about whether a child is safe, safe with intervention, or unsafe. This judgment was dichotomized into safe (safe and safe with intervention) and unsafe. In the case of the Smith family with the young child, 66.3% of the workers indicated that ''one or more safety threats are present and placement is the only protecting intervention possible for the children'' and 33.7% of workers did not have this finding. In the case of the Samuels family, 94.7% of workers did not find the child unsafe based on the above criteria and 5.3% of workers did decide that the child was unsafe. The OFRA has a possible abuse score that ranges from 0 to 18. The highest score given to the Smith child was 15 and the lowest was 1 with a mean of 4.82 ($SD = 2.14$). Scores clustered around a rating of 4, 5, or 6 with 62.4% of respondents rating the risk in that range. The highest score given to the Samuels child on the ORAM was 7 and the lowest was 1 with a mean of 3.4 ($SD = 1.24$). Scores were generally 3 or 4 with 66.6% of scores in this range. The ORAM has 5 levels of risk. A total of 8.5% of workers found the Smith child to be at low or no risk, 34% found the child to be at intermediate risk, 41.5% found the child to be at moderately high risk, and 16% found the child to be at high risk. For the Samuels child, 24% of workers found the child to be at low or no risk, 38.5% found the child to be at moderate risk, 26% found the child to be at moderately high risk, and only 1 worker found the child to be at high risk.

Worker's level of education and age were not associated with scores on the risk assessment measures. Risk assessment scores between the Smith and Samuels families were associated ($p \leq .001$), indicating that workers who attributed higher risk in one family also tended to attribute high risk in the other family.

Confidence in Performance and Assessment

Confidence in performance and confidence in the assessment of risk was positively associated with age. That is, as age increased, confidence also increased. There were no significant differences in confidence in either performance or confidence in assessment of risk based on gender. Furthermore, confidence in performance and risk were positively associated with one another in both scenarios. Thus, as a general rule, participants were confident in both their performance and judgment across scenarios (see Table 1).

Confidence in performance and confidence in the assessment of risk was negatively associated with acute stress at the end of the scenario (see Table 1). Confidence was not related to whether the client was or was not confrontational in *t* test analyses ($p = .157$ to $p = .926$). Furthermore, confidence in performance or risk was unrelated to the appraised level of risk for the child on any of the three standardized risk assessment measures using *t* test and Spearman's Rho analyses (OSA: $p = .432$ to $p = .534$; ORAM: $p = .570$ to $p = .558$; OFRA: $p = .469$ to $p = .550$).

Subjective Factors Affecting Confidence

The qualitative interviews were analyzed to examine statements related to high (Level 5) and low (Level 1 or 2) worker confidence in both their performance and the final risk assessment level assigned. In total, 9 participants rated their confidence in their performance as low (a score of 1 or 2) on the first scenario and 13 participants rated it low on the second scenario. As only one participant rated their confidence in performance as low in both scenarios, interviews of 21 different people were analyzed on this dimension. Nine participants rated their confidence in performance as high (a score of 5) on Scenarios 1 and 9 participants did so on Scenario 2. As 6 participants scored themselves as high on both scenarios, 12 individual interviews were analyzed on this dimension. Thirteen participants rated their confidence in risk assessment as low on Scenario 1 as did 10 on Scenario 2. There were three people who rated themselves low on both, resulting in interviews of 20 different people. Nine people rated their confidence in risk assessment as high in Scenarios 1 and 19 rated

Table 1. Correlations (Spearmans' Rho) Between Confidence in Performance and Confidence in Risk, Age, and Acute Stress Symptoms

	Confidence in Risk Scenario 1	Confidence in Performance Scenario 2	Confidence in Risk Scenario 2	Participant's Age	Acute Stress (STAI)
Confidence in Performance Scenario 1	.505***	.408***	.428***	.222*	−.439***
Confidence in Risk Scenario 1		.505***	.645***	.314**	−.287**
Confidence in Performance Scenario 2			.574***	.306**	−.294**
Confidence in Risk Scenario 2				.271**	−.275**

Note: STAI = State-Trait Anxiety Inventory.
*** $p \leq .001$.
** $p \leq .01$.
* $p \leq .05$.

it as high in Scenario 2, with 9 people scoring themselves as high in both.

High confidence in performance. Workers who felt confident in their performance felt that they used the time effectively and quickly collected data. Confident workers felt they had moved the interview along well and had been able to gain enough information to make an assessment of the child's risk of abuse.

Worker engagement with the client was also important in determining the level of confidence workers had in their performance. Workers who were confident felt they were able to build a rapport despite safety concerns and indicated that they remained calm despite the mother's emotional state. Confident workers indicated that they felt comfortable and did not feel threatened even when the mother was confrontational. For example, one worker indicated that she or he could overlook "jabs and remarks" made by a confrontational mother as the worker moved toward building rapport. Another indicated that she or he tried to accommodate the mother's mood and body language.

It was also noted by several participants that getting the mother's agreement or cooperation increased the sense of confidence in the worker regarding his or her own performance. Workers who felt confident in their performance were able to get the mother to acknowledge a willingness to work with child protective services and complete an assessment. In one case, it was noted that while the mother was resistant to child welfare involvement, she was willing to allow the worker to arrange community services. Thus, the ability to establish a safety plan through the mother's cooperation increased the worker's confidence in their own performance.

Low confidence in performance. Workers who felt low confidence in their performance indicated that they felt unprepared and disorganized. They felt they could not engage the mother or get her to answer questions effectively. They felt they did not ask the right questions or enough questions. One worker felt she or he was too critical. Two workers felt that they had made a mistake by stating the nature of the referral. One worker indicated that she or he is not sure what she or he would have done in the interview if there had been more time beyond the 15-min interview limit. Another worker indicated that she or he usually lacked confidence in initial interviews but that confidence increased in subsequent meetings with a client.

Although workers felt low confidence with both the confrontational and non-confrontational mother, workers indicated that they felt caught off guard, flustered, and uncomfortable. One worker indicated that "the control piece just died." Another stated that she or he could not get the situation under control. As a result of these feelings, one worker indicated that she or he could not remember the safety assessment questions. One worker indicated that she or he felt "really bad," scrambled, nervous, and "psychologically trapped for words." She or he wished that she or he was not alone and recommended that investigators should work in pairs to increase the sense of confidence. A worker indicated that she or he did not work well with hostile clients and thus would have liked a cool down period before completing the investigation. Another worker indicated that she or he "freezes up" with hostile clients. On the other hand, a worker with low confidence in his or her performance stated that confrontational parents are more the norm and thus having a non-confrontational "standoffish" parent threw him or her off.

High confidence in risk appraisal. Interestingly, as indicated above, confidence was not related to the appraised level of risk to the child, and workers felt equally confident whether they appraised the child to be at high risk or at low risk. When risk was judged to be high, workers that were confident in their risk appraisal focused on the seriousness of the abuse (such as the child getting hit with a belt or being burned) and the child's fear. Collateral information such as medical corroboration or the fact that there was a report from the day care provider was another factor that increased confidence. "Day cares do not lie about bruising."

The perception that the mother denied or minimized the abuse increased the workers' confidence that the child was at risk. One worker indicated that she or he was concerned by how well the mother could lie and thus she would not be trustworthy in the future. Another felt the mother was withholding information. The mother's lack of appreciation for her child's distress increased confidence that this was high risk. Workers felt swayed by the mother's emotional and psychological state. The mother's sense of isolation, lack of awareness of resources, and

lack of trust in the day care or school increased confidence that this was a high-risk situation.

When the risk was perceived to be low, the cooperation of the mother increased confidence in the appraisal, even when the mother was confrontational. One worker overlooked the confrontational behavior, noting that the confrontational mother took ownership and responsibility. Indeed at times being confrontational was considered a good sign. One worker stated she or he wished the mother was more confrontational, as this would demonstrate that she loved her daughter. Several workers felt that the mother's account of the injury to the older child was plausible (Ms. Samuels) and were confident that she did not cause the injury. In part, this was due to the fact that the child did not disclose abuse to the worker and in part due to the fact that the mother cared for the child's physical needs on other occasions. Furthermore, these workers did not feel that the mother would blame or punish the child for what the worker believed to be a false allegation. With respect to the younger child (Ms. Smith), workers who felt that the child was at low risk believed that the day care and school support was protective. It was stated that mother agreed to follow up contacts with community supports. Furthermore, the mother's willingness to have the worker contact mother's family increased confidence that the child was safe. Several workers also noted that the fact that the mother was employed increased their confidence that the child was not at risk.

Workers attributed high levels of confidence in the level of risk assessed to training, past supervisions, and experience. One worker indicated the tools and guidelines helped her or him make decisions. Another indicated that she or he was very familiar with the tools and therefore had high confidence. Another indicated that she or he had experience on working with family dynamics and grief responses.

Low confidence in risk appraisal. Workers with low confidence in their risk appraisal felt that they were unable to obtain sufficient information in their interview with the mother. Some workers attributed this to time restrictions, others attributed it to the mother's resistance to providing information, others attributed this to their own inability to engage the mother, and one because of lack of experience. One worker stated this was because she or he was too passive in the interview. A worker indicated that she or he became distracted by the mother's emotional state and behavior and did not ask sufficient questions to determine risk.

Discrepancies in the story between the mother and the reported abuse decreased confidence in risk assessment appraisal, leading one worker to "feel confused." Workers with low confidence when faced with discrepancies felt they could not determine whether the mother was lying or not. One worker with low confidence felt that the mother was open and honest about the areas where she needed support but remained adamant that she did not perpetrate abuse. As a result, the worker could not be sure about the level of risk. Another worker felt positively toward the mother and had difficulty reconciling this with the severity of the abuse. One worker believed that the child was not safe but did not feel she or he had sufficient information to support this.

Some workers with low confidence in the risk appraisal were concerned about the mother's lack of family support and her unstable relationship with her boyfriend. They were concerned about the stability of the mother's work, poverty, and lack of community resources and were also worried about the seriousness of the struggles that the mother was having in her own life, but were not sure if this translated into risk to the child. As a result, some workers stated that they were unable to determine the immediate safety risks.

One worker suggested that a further reason for low confidence in the risk appraisal is that the risk assessment tools are not robust. Another worker indicated that she or he had different findings of risk on the three assessment tools and therefore this decreased the level of confidence.

Discussion and Conclusions

In an effort to improve the ability of child welfare agencies and individual workers within these agencies to accurately identify children at risk of harm, child welfare services have moved toward highly standardized risk assessment models. While research in a variety of fields has demonstrated that carefully validated actuarial models outperform clinical judgment in estimating future risk (Dawes, 1994; Grove & Meehl, 1996), clinical judgment is nevertheless understood to remain critical to the complex process of predicting risk (Meehl, 1954; Schwalbe, 2008; Van de Luitgaarden, 2009). The results of this study confirm that even when presented with the same two families, child welfare workers are highly variable in their assessment of risk with some workers viewing risk of each child to be high and others viewing it to be low. Despite the use of validated measures, they assess risk differently.

In addition to examining the consistency with which workers make judgments of risk, this research sought to understand the confidence with which workers made assessments of risk and subjective factors that influence both confidence and judgment. Previous research has indicated that professional judgment and confidence are both related to experience (Gambrill, 2005; Hay et al., 2008). This study similarly found that age was related to confidence in both performance and risk assessment. The findings were evident not only in the quantitative analysis but also in the qualitative interviews where workers discussed their extent and quality of their training, experience and past supervision as contributing to confidence regarding risk assessment. Worker confidence tended to be consistent both across the dimensions of performance and assessment of risk and across scenarios, suggesting that confidence in judgment and clinical ability is relatively stable. In addition, workers who had lower levels of confidence experienced higher subjective levels of anxiety. However, confidence was not related to the level of assessed risk. That is, workers could be equally confident in higher and lower attributions of risk. Furthermore, age and educational level were not associated with assessed level of risk. Previous research confirms that

experience does not explain differences in risk assessment judgments (Arad-Davidzon & Benbenishty, 2008; Gold, Benbenishty, & Osmo, 2001).

Shlonsky and Wagner (2005) propose that actuarial measures do not assist workers to develop a cooperative relationship with the family in terms of risk planning. Indeed, in this study, workers felt that their ability to engage the parents in the case scenarios influenced both their confidence in their performance and their confidence in assessing the child's level of risk. Those workers were able to engage the parent in cooperative planning for the child and, further, confident workers reported that they were unmoved by confrontational behavior on the part of the parent. In fact, some workers expected confrontation, viewing it positively and expressing concern when it was absent. On the contrary and consistent with previous research (Horwath, 2006), workers who did not express confidence reported a further eroding of confidence when the parent was confrontational. Such workers appear to have experienced the confrontational nature of the encounter as overwhelming and distracting, which led them to freeze or fragment. These emotional states led them to comment that they could not stay focused or ask the appropriate questions nor could they engage the mother. This is similar to previous research (Van de Luitgaarden, 2009) that found workers who were less confident were unable to acquire the necessary information to make a judgment of risk. Other case characteristics were also associated with increased confidence regarding risk assessment. These factors included the mother's willingness to share information, the mother's support networks, and the mother's appreciation of the child's distress. Workers' confidence in their performance and in their risk appraisal were affected by their interpretation of objective facts such as the seriousness of the abuse and collateral information, as well as the mother's behaviors such as being withholding, inability to appreciate the child's distress, her isolation, unfamiliarity with resources, and mistrust of school or day care.

A limitation of this study was the short time frame in which assessments needed to be completed (15 min). Interestingly, however, the limited time frame still allowed for many workers to make a decision about risk with confidence. This is consistent with decision-making theory and research in which clinicians have been found to form judgments early in the assessment process that are then confirmed by selective attention to subsequent information. In part, this can be due to a reliance on the disposition of the client as representative of their future behavior or by a belief that the client is representative of other cases the clinician has encountered (Arkes, 1981; Baumann et al., 1991; Glascoe & Dworkin, 1993; Gambrill & Shlonsky, 2000). In this way, previous experiences sway current judgments. A further limitation is the fact that all scenarios involved White mothers. This decision was made in order to control the number of variables that may influence judgment and due to sample size limitations. However, given the importance of these variables in previous research, this would certainly be an area for further investigation.

Assessment of risk in child welfare could be considered a classic example of micro-certainty with macro-uncertainty (Baumann et al., 1991). That is, predicting which children are at risk is a near impossible task that child protection workers heroically undertake, despite difficult odds. The variation in risk assessment appraisals in this study despite at times high rates of confidence speaks to the need for ongoing consultation, increased attention to more subjective elements of risk assessment tools, and more focus on critical thinking skill especially in relation to individual heuristic strategies even among highly skilled and trained workers. Furthermore, training in engagement and assessment processes in the context of uncertainty as well as an admission by policy makers and managers with respect to the very real limitations of risk assessment tools remain constants for the field. Despite improved decision-making capacity through the use of reliable and valid risk assessment tools, the risk of error is simply too high to fail to invest in improving the ongoing clinical skills of child protection workers.

Declaration of Conflicting Interests

The author(s) declared no conflicts of interests with respect to the authorship and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research and/or authorship of this article: grant from the Social Sciences and Humanities Research Council of Canada.

References

Arad-Davidzon, B., & Benbenishty, R. (2008). The role of worker's attitudes and parent and child wishes in child protection workers' assessments and recommendations regarding removal and reunification. *Children and Youth Services Review*, *30*, 107-121.

Arkes, H. (1981). Impediments to accurate clinical judgment and possible ways to minimize their impact. *Journal of Counselling and Clinical Psychology*, *49*, 232-330.

Baird, C., Wagner, D., Healy, T., & Johnson, K. (2000). The relative validity of actuarial- and consensus-based risk assessment systems. *Children and Youth Services Review*, *22*, 839-871.

Barber, J., Shlonsky, A., Black, T., Goodman, D., & & Trocmé, N (2008). Reliability and predictive validity of a consensus-based risk assessment tool. *Journal of Public Child Welfare*, *2*, 173-195.

Baumann, A., Deber, R., & Thompson, G. (1991). Overconfidence among physicians and nurses: The micro-certainly, macro uncertainty phenomenon. *Social Science Medicine*, *32*, 167-174.

Bogo, M., Regehr, C., Woodford, M., Hughes, J., Power, R., & Regehr, G. (2006). Beyond competencies: Field instructors' descriptions of student performance. *Journal of Social Work Education*, *42*, 191-205.

Camasso, M. J., & Jagannathan, R. (2000). Modeling the reliability and predictive validity of risk assessment in child welfare. *Children and Youth Services Review*, *22*, 873-896.

Chu, W., & Tsui, M. (2008). The nature of practice wisdom in social work revisited. *International Social Work*, *51*, 47-54.

Cicchinelli, L. F. (1991). *Proceedings from the Symposium on Risk Assessment in Child Protective Services*. National Center on Child Abuse and Neglect, Washington, DC.

Dawes, R. M. (1994). *House of cards: Psychology and psychotherapy built on myth.* New York, NY: Free Press.

DePanfilis, D., & Zuravin, S. J. (2001). Assessing risk to determine need for services. *Children and Youth Services Review*, *23*, 3-20.

English, D. J., & Pecora, P. J. (1994). Risk assessment as a practice in child protective services. *Child Welfare*, *73*, 451-473.

Fluke, J. D., Baumann, D., England, P., Kern, H., Law, R., McFadden, T., & Schultz, F. (1995). Emerging critical conceptual issues in risk assessment research and practice. In T. Tatara (Ed.), *Eighth National Roundtable on CPS risk assessment* (pp. 239-250). Washington, DC: American Public Welfare Association.

Gambrill, E. (2005). Decision making in child welfare: Errors and their context. *Children and Youth Services Review*, *27*, 347-352.

Gambrill, E., & Shlonsky, A. (2000). Risk assessment in context. *Children and Youth Services Review*, *22*, 813-837.

Gambrill, E. (2006). *Social work practice: A critical thinkers guide* (2nd ed.). New York, NY: Oxford University Press.

Gibbs, L. (2003). *Evidence-based practice for the helping professions: A practical guide with integrated multimedia.* Pacific Grove, CA: Brooks/Cole-Thomson Learning.

Glascoe, F., & Dworkin, P. (1993). Obstacles to effective developmental surveillance: Errors in clinical reasoning. *Developmental and Behavioral Pediatrics*, *14*, 344-349.

Gold, N., Benbenishty, R., & Osmo, J. (2001). A comparative study of risk assessment and recommended interventions in Israel and Canada. *Child Abuse and Neglect*, *25*, 409-429.

Goldberg, L., & Werts, C. (1966). The reliability of clinicians judgments: A multitrait-multimethod approach. *Journal of Consulting Psychology*, *30*, 199-206.

Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy, and Law*, *2*, 293-323.

Hay, M., Weisner, T., Subramanian, S., Duan, N., Niedzinski, E., & Kravitz, R. (2008). Harnessing experience: Exploring the gap between evidence-based medicine and clinical practice. *Journal of Evaluation in Clinical Practice*, *14*, 707-713.

Haynes, R. B., Devereaux, P. J., & Guyatt, G. H. (2002). Clinical expertise in the era of evidence-based medicine and patient choice [Editorial]. *ACP Journal Club*, March/April, *136*, 1–7.

Horwath, J. (2006). The missing assessment domain: Personal, professional and organizational factors influencing professional judgments when identifying and referring child neglect. *British Journal of Social Work*, *37*, 1285-1303.

Johnson, W., & L'Esperance, J. (1984). Predicting the recurrence of child abuse. *Social Work Research and Abstracts*, *20*, 21-26.

Kane, M. T. (1992). The assessment of professional competence. *Evaluation & The Health Professions*, *15*, 163-182.

Laughlin, M. (2008). Reason, reality and objectivity—share dogmas and distortions in the way both "scientistic" and "postmodern" commentators frame the EBM debate. *Journal of Evaluation in Clinical Practice*, *14*, 665-671.

Lindsey, D. (1992). Reliability of the foster care placement decision: A review. *Research on Social Work Practice*, *2*, 65-80.

Lyons, P., Doueck, H. J., & Wodarski, J. S. (1996). Risk assessment for child protective services: A review of the empirical literature on instrument performance. *Social Work Research*, *20*, 143-154.

Mamassian, P. (2008). Overconfidence in an objective anticipatory motor task. *Psychological Science*, *19*, 601-606.

McCracken, S., & Marsh, J. (2008). Practitioner expertise in evidence-based practice decision making. *Research on Social Work Practice*, *18*, 301-310.

McDonald, T., & Marks, J. (1991). A review of risk factors assessed in child protective services. *Social Service Review*, *65*, 112-132.

Meehl, R. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence.* Minneapolis: University of Minnesota Press.

Millar, K., Jelicic, M., Bonke, B., & Asbury, A. (1995). Assessment of preoperative anxiety: comparison of measures in patients awaiting surgery for breast cancer. *British Journal of Anaesthesia*, *74*, 180-183.

Miller, J., Williams, K., English, D., & Olmstead, J. (1987). Risk assessment in child protection: A review of the literature. *Occasional monograph series of APWA Social R&D Department (No. 1).* Washington, DC: American Public Welfare Association.

Moerman, N., van Dam, F., Muller, M., & Oosting, H. (1996). The Amsterdam Preoperative Anxiety and Information Scale (APAIS). *Anesthesia & Analgesia*, *182*, 445-451.

Pecora, P. J. (1991). Investigating allegations of child maltreatment: The strengths and limitations of current risk assessment systems. In M. Robin (Ed.), *Assessing child maltreatment reports: the problem of false allegations* (pp. 73-92). New York, NY: Haworth Press.

Sackett, D. L., Straus, S. E., Richardson, W. S., Rosenberg, W., & Haynes, R. B. (2000). *Evidence-based medicine: How to practice and teach EBM* (2nd ed.). New York, NY: Churchill Livingstone.

Schwalbe, C. (2008). Strengthening the integration of actuarial risk assessment with clinical judgment in an evidence based practice framework. *Children and Youth Services Review*, *30*, 1458-1464.

Shlonsky, A., & Wagner, D. (2005). The next step: Integrating actuarial risk assessment and clinical judgment into an evidence-based practice framework in CPS case management. *Children and Youth Services Review*, *27*, 409-427.

Smith, D., & Dumont, F. (2002). Confidence in psychodiagnosis: What makes us so sure? *Clinical Psychology and Psychotherapy*, *9*, 292-298.

Spielberger, C. (1983). *Manual for the State-Trait Anxiety Inventory.* Palo Alto, CA: Consulting Psychologists Press.

Van de Luitgaarden, G. (2009). Evidence-based practice in social work: Lessons from judgment and decision-making theory. *British Journal of Social Work*, *39*, 243-260.

Wagner, D., & Johnson, K. (2006). *The California structured decision making risk revalidation: A prospective study.* Madison, WI: Children's Research Center.

Wald, M. S., & Woolverton, M. (1990). Risk assessment: The emperor's new clothes? *Child Welfare*, *69*, 483-511.