

^a Uniwersytet w Hajfie, Izrael

^b Wydział Psychologii Społecznej i Rozwojowej, Uniwersytet Cambridge,
Free School Lane, Cambridge, CB2 3RQ, Wielka Brytania

^c Ministerstwo Pracy i Polityki Społecznej, Jerozolima, Izrael

Poprawa oceny wiarygodności oskarżeń w przypadkach wykorzystywania seksualnego dzieci – rola protokołu przesłuchania dochodzeniowego wg NICHD

Cel: Celem badania było ustalenie, czy wiarygodność zeznań dzieci dotyczących domniemanych incydentów wykorzystywania seksualnego można oceniać w sposób bardziej rzetelny i niezawodny, jeśli przesłuchanie dochodzeniowe przebiega według protokołu NICHD, a nie w sposób improwizowany.

Metody: Grupa czterdziestu dwóch izraelskich specjalistów dochodzeniowych, posiadających stosowne doświadczenie w pracy z młodzieżą oceniała wiarygodność oskarżeń o wykorzystywanie seksualne sformułowanych przez domniemane ofiary takiego wykorzystywania w toku przesłuchania prowadzonego z uwzględnieniem protokołu lub bez niego. W połowie przypadków na podstawie niezależnego materiału dowodowego zajście domniemanego zdarzenia uznano za przekonujące („prawdopodobne”), drugą połowę uznano za mało prawdopodobną („nieprawdopodobne”).

Wyniki: Stwierdzenie „Jednoznaczna ocena niemożliwa” częściej odnosiło się do przesłuchań nieuwzględniających protokołu niż tych przebiegających zgodnie z protokołem. W porównaniu z oskarżeniami sformułowanymi w czasie przesłuchań bez protokołu, prawdopodobne i nieprawdopodobne zarzuty stawiane w toku przesłuchań wg protokołu były bardziej trafnie oceniane jako, odpowiednio, wiarygodne lub niewiarygodne. Dla przesłuchań przebiegających zgodnie z protokołem wyższy był też poziom zgodności między niezależnymi oceniającymi (inter-rater reliability). Różnice były jednak istotne tylko w grupie przypadków prawdopodobnych.

Wnioski: Wykorzystanie protokołu przesłuchania dochodzeniowego wg NICHD ułatwia przesłuchującemu ocenę wiarygodności zeznania, choć zarzuty mało wiarygodne (opisujące zdarzenia, które prawdopodobnie nie miały miejsca) pozostają trudne do wykrycia, nawet z zastosowaniem protokołu.

Wprowadzenie

Lepsze zrozumienie zeznań dzieci dotyczących doświadczonego wykorzystywania to przedmiot aktywności wielu badaczy. Ich wysiłki skupiają się na ogół na opracowaniu strategii przesłuchania pozwalającej uzyskać od dziecka większą ilość informacji lepszej jakości, ale kwestia oceny wiarygodności dzieci–świadków także nie pozostaje niezauważona. Choć dostępne są pewne narzędzia służące ocenie takiej wiarygodności, nigdy nie poddano ich walidacji w warunkach empirycznych. Mimo to wielu specjalistów deklaruje posiadanie umiejętności w tej dziedzinie, a decyzje dotyczące wiarygodności zeznań podejmowane są każdego dnia przez śledczych, pracowników socjalnych, sędziów i przysięgłych. Celem niniejszego badania było ustalenie, czy wiarygodność zeznania można oceniać w sposób bardziej trafny i rzetelny, jeśli przesłuchanie dochodzeniowe prowadzi doświadczony specjalista przestrzegający wymogów „dobrej praktyki”.

W ramach podobnych projektów badawczych — dotyczących jednak opisów zainscenizowanych zdarzeń, a nie dochodzenia w sprawie domniemanego przestępstwa — badacze opiniowali ocenę wiarygodności realizowaną przez specjalistów z wykorzystaniem analizy prawdziwości relacji (SVA — Statement Validity Analysis) i jej składowej — opartej na kryteriach analizie treści (CBCA — Criterion Based Content Analysis) (Akehurst, Koehnken, Hofer 2001; Steller, Wellerhaus, Wolf 1988; Yuille 1988) lub obserwacji zachowań pozawerbalnych w czasie przesłuchania dochodzeniowego (Ceci, Crotteau–Huffman 1997; Westcott, Davies, Clifford 1991). Inne badania skupiały się na ocenie wiarygodności w oczach nie–specjalistów (Lewis 1993; Lewis, Stanger, Sullivan 1989; Vrij, Van Wijngaarden 1994) lub na porównaniu ocen specjalistów i nie–specjalistów (Chahal, Cassidy 1995; Jackson 1996; Tye, Henderson, Honts 1995). Trafność, z jaką re-

lacje rozpoznawano jako prawdziwe lub fałszywe, była w tych badaniach różnicowana, na ogół jednak jedna trzecia ocen okazywała się błędna. Oceniający, którzy zdawali się przede wszystkim na analizę treści, poprawnie rozpoznawali relacje prawdziwe i fałszywe w 46% (Ceci, Crotteau–Huffman 1997) do 89% (Tye i wsp. 1995) przypadków. Z kolei oceniający, uwzględniający wskazówki pozawerbalne, prawidłowo interpretowali od 43% (Jackson 1996) do 68% (Chahal, Cassidy 1995) relacji. Wydaje się, że specjaliści nie potrafili często rozróżnić opisu faktycznie doświadczonych zdarzeń od zdarzeń, które nie miały miejsca (Ceci, Loftus, Leichtman, Bruck 1994; Horner, Guyer, Kalter 1993a, 1993b; Leichtman, Ceci 1995). Wyniki uzyskiwane przez profesjonalistów nie zawsze są lepsze niż te uzyskiwane przez nie–specjalistów lub studentów. Choć w jednym badaniu wykazano, że specjaliści radzili sobie lepiej niż laicy (Tye i wsp. 1995), doniesienia Ceci i Crotteau–Huffman (1997) nie potwierdzają takiego wniosku, a często okazuje się wręcz, że specjaliści mniej skutecznie niż nie–specjaliści rozróżniają relacje ze zdarzeń rzeczywistych i fikcyjnych (Ceci i wsp. 1994; Horner i wsp. 1993a, 1993b; Leichtman, Ceci 1995).

Ponieważ oceniający mają tak duże problemy z jednoznacznym określeniem, czy dane zeznanie jest prawdziwe czy fałszywe — niezależnie od tego, czy pochodzi ono z badań wykorzystujących inscenizowane zdarzenia (Ceci i wsp. 1994; Horner i wsp. 1993a, 1993b; Poole, Lindsay 1997), czy zaczerpnięto je z rzeczywistych przesłuchań dochodzeniowych (Finlayson, Koocher 1991; Jackson, Nuttal 1993) — poziom zgodności między dwoma niezależnymi oceniającymi jest także niepokojąco niski. Rozkład ocen danego zeznania obejmuje na ogół cały dostępny zakres. Dla przykładu, zakres w badaniu Horner i wsp. sięgał od 0,10 do 0,90, a w badaniu Finlayson i Koocher od

0–25% do 75–100%. W pracy Jackson i Nuttal, oceny tego samego zapisu zeznania rozciągały się od stwierdzeń „z całą pewnością doszło do krzywdzenia” do „z całą pewnością nie doszło do krzywdzenia”. Choć wysoka zgodność między oceniającymi nie przekłada się bezpośrednio na prawidłowość opiniowania, niska zgodność oznacza, że przynajmniej część ocen jest niepoprawna.

Badania terenowe wykorzystujące metody takie jak CBCA donoszą o znacznie wyższej skuteczności odróżniania zarzutów wiarygodnych od niewiarygodnych (Boychuk 1991; Lamb i wsp. 1997; Raskin, Esplin 1991a, 1991b), ale krytycy poddają w wątpliwość procedury zastosowane w ramach rzeczonych projektów w celu określenia, czy dana relacja jest prawdziwa czy fałszywa (np. Wells, Loftus 1991). Niezmiernie trudno jest **udowodnić**, że domniemane zdarzenie nie miało miejsca: w pracy Lamb i wsp. (1997) w ramach przeglądu systematycznego ponad 1 000 spraw toczących się w całym Izraelu w okresie objętym badaniem na początku lat dziewięćdziesiątych udało się wyodrębnić zaledwie 13 takich niezaprzeczalnie „nieprawdopodobnych” zarzutów. Analogicznie wyodrębnione zarzuty „prawdopodobne” (tzn. te o „dowodzonej” prawdziwości) uzyskiwały istotnie wyższe wyniki w ocenie CBCA niż zarzuty „nieprawdopodobne”, ale różnice nie były duże, a błędne interpretacje na tyle częste, że trudno mówić o możliwości zastosowania jej w praktyce sądowej. Hershkowitz, Lamb, Sternberg i Esplin (1997) wykazali później, że wyodrębnienie kryteriów do CBCA w wypowiedziach dzieci było łatwiejsze, gdy przesłuchującym udawało się uzyskiwać od domniemanych ofiar odpowiedzi narracyjne. Niestety badania wskazują, że śledczy rzadko usiłują sprowokować wypowiedzi narracyjne (w USA: Craig, Sheibe, Kircher, Raskin, Dodd 1999; Sternberg, Lamb, Hershkowitz 1996; w Izraelu: Lamb i wsp. 1996; w Szwecji: Cederborg, Orbach, Sternberg, Lamb 2000; w Wielkiej Brytanii:

Sternberg, Lamb, Davies, Westcott 2001), o ile nie wykorzystują protokołu przesłuchania śledczego według NICHD (Orbach i wsp. 2000; Sternberg, Lamb, Orbach, Esplin, Mitchell 2001).

Protokół NICHD obejmuje wszystkie fazy przesłuchania dochodzeniowego. Sens jego powstania polega na przekładaniu zaleceń wynikających z badań na wytyczne praktycznego postępowania umożliwiającego pozyskiwanie konkretnych, pełnych i precyzyjnych relacji dotyczących krzywdzenia od domniemanych małoletnich ofiar i świadków. Wypowiedziom narracyjnym ma sprzyjać wspierająca atmosfera przesłuchania (wstępne budowanie relacji), zastosowanie strategii przesłuchania do poziomu rozwojowego dziecka i jego możliwości (np. obniżenie złożoności językowej i unikanie wtrąceń), przygotowanie dziecka do roli związanej z przekazywaniem informacji (wyjaśnienie zasad komunikacji, nauka relacjonowania wspomnień epizodycznych odnoszących się do zdarzenia) i maksymalne wykorzystywanie przez przesłuchującego wypowiedzi wymagających swobodnego odtwarzania. Przesłuchujący są w ramach protokołu instruowani, aby w możliwie największym zakresie korzystać z pytań otwartych, wprowadzać pytania szczegółowe dopiero po wyczerpaniu możliwości strategii otwartych, pytania oparte na wyborze opcji wprowadzać dopiero pod koniec przesłuchania w celu uzyskania kluczowych informacji i rezygnować z praktyk sugerujących odpowiedź. Poprzez zastosowanie „pytań wprowadzających”, „segmentacji czasowej” i techniki „parowania” (patrz poniżej) przesłuchujący ma wykorzystywać informacje pozyskane już od dziecka, aby stymulować dalsze swobodne odtwarzanie. Przestrzeganie protokołu NICHD w czasie przesłuchania dzieci w wieku od 4 do 13 lat pozwala uzyskać około 50% informacji w trybie pytań otwartych, podczas gdy odsetek ten wynosi 10–20% jeśli przesłuchujący z proto-

kołu nie korzystają (Lamb i wsp. 2003; Orbach i wsp. 2000; Sternberg, Lamb, Davies i wsp. 2001; Sternberg, Lamb, Orbach i wsp. 2001). Jednym z celów niniejszego badania było ustalenie, czy profesjonalści opiniu-

jący prawdziwość zeznań dzieci rzeczywicie bardziej trafnie oceniali wiarygodność relacji jeśli pochodziła ona z przesłuchania zgodnego z protokołem niż w przypadku relacji z przesłuchania standardowego.

Ryc. 1. Protokół przesłuchania dziecka wg NICHD*

Protokół przesłuchania opracowany przez specjalistów NICHD (amerykańskiego Narodowego Instytutu Rozwoju i Zdrowia Dziecka) ma zapewnić dziecku możliwość przećwiczenia szczegółowego relacjonowania doświadczonych zdarzeń zanim przesłuchujący zapyta o zdarzenia będące przedmiotem dochodzenia. Procedura NICHD nakłada na przesłuchującego obowiązek pouczenia dziecka o konieczności mówienia prawdy oraz zachęcania go do poprawiania przesłuchującego i mówienia „nie wiem” tam, gdzie to konieczne. Protokół przesłuchania wg NICHD składa się, w kolejności, z następujących faz lub etapów:

1. Przedstawienie uczestników przesłuchania i ich ról
2. Omówienie definicji prawdy i kłamstwa
3. Budowanie relacji i ocena języka
4. Opis niedawnego istotnego wydarzenia
5. Wstępna swobodna relacja dotycząca oskarżenia
6. Relacja z ostatniego zdarzenia (jeśli dziecko opisuje liczne zdarzenia)
7. Pytania wspomagające (np. „Mówiłeś coś o stodole. Opowiedz mi o tym.”)
8. Pytanie wprost o ostatnie zdarzenie uzupełnione pytaniem otwartym
9. Relacja z pierwszego zdarzenia
10. Pytania wspomagające
11. Pytanie wprost o pierwsze zdarzenie uzupełnione pytaniem otwartym
12. Relacje dotyczące innego zdarzenia, dobrze pamiętanego przez dziecko
13. Pytania wspomagające
14. Ułożone w pary pytania bezpośrednio–otwarte dotyczące tego zdarzenia
15. Jeśli to konieczne, pytania naprowadzające dotyczące szczegółów ważnych z punktu widzenia dochodzenia, a niewspomnianych przez dziecko
16. Zachęta do podzielenia się innymi dowolnymi informacjami, o których dziecko chce porozmawiać
17. Powrót do neutralnego tematu

* Rycina ta nie występuje w oryginalnym tekście artykułu, została dodana przez Redakcję.

W ramach licznych badań eksperymentalnych dzieci proszone są o opowiadanie nieprawdziwych historii dotyczących zdarzeń mało istotnych i nieznaczących emocjonalnie. Uogólnianie uzyskanych w ten sposób wyników i przenoszenie ich na ocenę wiarygodności w kontekście procesu dochodzeniowego wydaje się więc dość problematyczne, mimo że badania takie uwzględniają często sugerujące i powtarzające się pytania dotyczące kontaktu cielesnego, a badacze unikają wprowadzania informacji niepod-

nych wcześniej przez dziecko (Ceci, Crotteau–Huffman 1997; Leichtman Ceci 1995). Z tego względu niniejsze badanie objęło wiarygodne i niewiarygodne zarzuty dotyczące wykorzystywania seksualnego, stawiane przez dzieci w toku przesłuchań dochodzeniowych prowadzonych w Izraelu w latach 1994–2001 przez profesjonalnych śledczych wyspecjalizowanych w pracy z młodzieżą, a od 1998 roku zobowiązanych do wykorzystywania protokołu NICHD. Połowa analizowanych przesłuchań miała miejsce przed

wprowadzeniem obowiązku stosowania protokołu, a druga połowa — zrealizowana przy udziale tych samych przesłuchujących — po nałożeniu takiego obowiązku.

Zarzuty były niezależnie opiniowane przez trzech oceniających jako prawdopodobne lub nieprawdopodobne. W ocenie tej opierano się na twardych dowodach (metoda „ground truth”), tj. przyznanie się sprawcy do winy, raporty świadków, dowody fizyczne (np. materiał fotograficzny) i wyniki badań medycznych (więcej informacji o ocenie prawdopodobieństwa na podstawie niezależnych dowodów odnoszących się do możliwości, że dane zdarzenie miało miejsce, patrz Horowitz i wsp. 1997; Lamb i wsp. 1997). Przypadki prawdopodobne zostały stosowanie zaklasyfikowane i włączone do badania, o ile dostępne były wystarczające dowody w formie zeznań niezainteresowanych świadków, którzy widzieli część lub całość zdarzenia, dowodów fizycznych (np. materiał fotograficzny lub wideo), wyników badań medycznych i/lub zeznań podejrzanego, zawierających szczegółowy opis zdarzenia zgodny z relacją domniemanej ofiary, dzięki czemu oceniający mógł z dużą dozą pewności stwierdzić, że do domniemanego zdarzenia rzeczywiście doszło. Relacje dotyczące tego rodzaju zdarzeń określane są w niniejszej pracy jako „wiarygodne”. Przypadki nieprawdopodobne zostały stosownie zaklasyfikowane i włączone do badania, o ile dostępne były podobne wystarczające dowody wskazujące, że do opisywanego zdarzenia nie doszło. Zarzuty były więc na przykład uznawane za nieprawdopodobne, jeśli niezainteresowany świadek widział domniemaną ofiarę i/lub podejrzanego we wskazanym momencie i nie zaobserwował oznak rzekomego zdarzenia; wyniki badań medycznych nie potwierdziły obecności urazów, których należałoby oczekiwać wskutek domniemanego zdarzenia i/lub obecne są fizyczne dowody wskazujące, że do domniemanego zdarzenia nie doszło. W niektórych przypad-

kach domniemana ofiara wycofywała swoje oskarżenia w toku dalszych przesłuchań, przedstawiając równocześnie przekonującą alternatywną relację, znajdującą poparcie w niezależnych dowodach. W ramach niniejszej pracy relacje odnoszące się do zdarzeń, które najprawdopodobniej nie miały miejsca określane są jako „niewiarygodne”.

Połowę oskarżeń prawdopodobnych (zeznań wiarygodnych) i połowę oskarżeń nieprawdopodobnych (zeznań niewiarygodnych) postawiono w toku przesłuchań prowadzonych z użyciem protokołu przesłuchania dochodzeniowego NICHHD. Do wybranych przypadków dopasowano sprawy badane przez tego samego przesłuchującego przed wprowadzeniem obowiązku korzystania z protokołu. Dopasowanie przypadków uwzględniało wiek dziecka, rodzaj oskarżeń i moc dostępnych dowodów, zanim odszukano stosowne zapisy przesłuchań.

Ponieważ protokół NICHHD pozwala uzyskać większą ilość wypowiedzi narracyjnych opisujących domniemane zdarzenie niż ma to zazwyczaj miejsce w czasie przesłuchania dowodowego, przesłuchujący powinien móc uzyskać szerszą wiedzę o naturze zarzutów. Założyliśmy więc, że ocena wiarygodności oskarżeń postawionych w toku przesłuchania zgodnego z protokołem powinna być trafniejsza niż w przypadku przesłuchań nieuwzględniających protokołu.

Zarówno profesjonalści, jak i studenci z większą trafnością rozpoznawali relacje prawdziwe niż fałszywe (Akehurst i wsp. 2001; Steller i wsp. 1988; Tye i wsp. 1995; Yuille 1988). Przyczyną takiego stanu rzeczy może być fakt, że oceniający wykorzystują inne techniki w identyfikacji relacji prawdziwych niż fałszywych. W niektórych badaniach na przykład wykazano, że do wyodrębniania stwierdzeń prawdziwych w większym stopniu służyły informacje werbalne, podczas gdy materiał pozawerbalny służył przede wszystkim do rozpoznawania kłamstw (Anderson, De Paulo, Ansfield, Tickle, Green

1999; Vrij, Edward, Bull 2001). Jeśli wskazówki niewerbalne były w ocenie niedostępne, a opiniowanie opierało się wyłącznie na zapisach przesłuchań, w ocenie wiarygodności zdawało się przeważać domniemanie szczerości („*truth bias*”) (Akehurst i wsp. 2001; Steller i wsp. 1988; Tye i wsp. 1995; Yuille 1988). Oceniający wykazują też więcej pewności identyfikując relacje prawdziwe niż fałszywe (Anderson, De Paulo, Ansfield 1999; De Paulo, Charlton, Cooper, Lindsay, Muhlenbruck 1997; Vrij, Baxter 2000; Vrij, Harden, Terry, Edward, Bull 2001). W ramach niniejszego badania staraliśmy się określić, czy oce-

niający łatwiej, trafniej i z większą pewnością przypisywali wiarygodność relacjom prawdopodobnym niż brak wiarygodności relacjom fałszywym.

W ogólnym ujęciu przewidywalimy, że ocenawiarogodności będzie trafniejsza i pewniejsza, a poziom zgodności opiniujących będzie wyższy w przypadku przesłuchań prowadzonych z użyciem protokołu NICHHD, w porównaniu z brakiem protokołu. Oczekiwaliśmy też, że oceniający bardziej trafnie i z większą pewnością określą jako wiarygodne relacje „prawdopodobne” niż jako niewiarygodne relacje „nieprawdopodobne”.

Metody

Uczestnicy

W badaniu uczestniczyli wszyscy izraelscy przesłuchujący wyspecjalizowani w pracy z małoletnimi, którzy w okresie zbierania danych w 2003 roku byli aktywni zawodowo (N=42). Średnia wieku przesłuchujących wynosiła 34,5 roku (SD [odchylenie standardowe] = 10,1), a średnia lat doświadczenia zawodowego 4,6 roku (SD = 5,70). Prze-

słuchujący przeprowadzili średnio 678,6 (SD = 759,70) przesłuchań, w tym 150,5 (SD = 110,20) w roku poprzedzającym badanie. Każdy z nich korzystał średnio miesięcznie z 6,10 (SD = 5,70) godziny superwizji i szkolenia w dziedzinie praktyki przesłuchań. Żaden z uczestników nie był zaznajomiony z przypadkami objętymi badaniem.

Wybór i klasyfikacja przesłuchań

Przesłuchania oceniane w ramach badania, w szczególności te zawierające zarzuty uznawane za nieprawdopodobne, wybrano z dużej puli przesłuchań, biorąc pod uwagę istnienie jednoznacznego, niezależnego materiału dowodowego potwierdzającego prawdopodobieństwo lub nieprawdopodobieństwo zdarzeń. W celu oceny prawdopodobieństwa, relacje poddawano najpierw niezależnej ocenie trzech ekspertów wykorzystujących dowody zewnętrzne, takie jak przyznanie się sprawcy do winy, zeznania świadków, ustalenia medyczne i inne dowody fizyczne. Materiał dowodowy analizowa-

no z wykorzystaniem tzw. skali „*ground truth*” opracowanej przez Horowitz, Lamb, Esplin, Reiter-Lavery, Krispin (1995), a następnie zmodyfikowanej i wykorzystanej w ramach pracy Lamb i wsp. (1997). W niniejszym badaniu zgodność pomiędzy niezależnymi oceniaczami wynosiła 91% (kappa = 0,88). Rozbieżności poddawano dyskusji do osiągnięcia konsensusu. Przy pomocy skali „*ground truth*” zarzuty klasyfikowano jako „bardzo możliwe”, „możliwe”, „niemożliwe” lub „absolutnie niemożliwe”. Za przykładem badania Lamb i wsp. (1997), zarzuty ocenione jako „bardzo możliwe” i „możliwe” na po-

trzeby badania uznawano za prawdopodobne, a relacje „absolutnie niemożliwe” i „niemożliwe” za nieprawdopodobne.

Postępując w ten sposób wyodrębniono dwadzieścia cztery przesłuchania dochodzeniowe dzieci stawiających zarzuty wykorzystywania seksualnego, co do których niezależne materiały dowodowe pozwalały je uznać za prawdopodobne (12) lub nieprawdopodobne (12). Zapisy tych przesłuchań przekazano do oceny uczestnikom badania. W sześciu przypadkach przesłuchania prowadzone według protokołu zawierały zarzuty prawdopodobne; w sześciu sprawach przesłuchania według protokołu prowadziły do zarzutów nieprawdopodobnych; sześć kolejnych przesłuchań prowadzonych bez uwzględnienia protokołu zawierało zarzuty prawdopodobne, a ostatnie sześć przesłuchań nie uwzględniających protokołu prowadziło do zarzutów nieprawdopodobnych. Oskarżenia były w największym możliwym zakresie dopasowane pod względem wieku dzieci i ciężkości domniemanej krzywdy (dotyk przez ubranie, dotykanie pod ubraniami i penetracja). Każdy z uczestniczących specjalistów został poproszony o zaopiniowanie czterech zapisów, każdy z innej kategorii.

Ocena wiarygodności. Siedmiu specjalistów dochodzeniowych niezależnie oceniali wiarygodność poszczególnych zapisów przesłuchań. Uczestnicy nie mieli dostępu do materiału dowodowego wykorzystanego w ramach metody „ground truth”, oceniali więc relacje wyłącznie na podstawie dostarczonych zapisów przesłuchania. Ocena polegała na określeniu prawdopodobieństwa, że do domniemanego zdarzenia faktycznie doszło, z wykorzystaniem 4-stopniowej skali: „bardzo możliwe”, „całkiem możliwe”, „raczej niemożliwe” lub „absolutnie niemożliwe”. Do dyspozycji była także opcja „Jednoznaczna ocena niemożliwa” (JON). Uczestnicy dodatkowo wskazywali na 5-stopniowej skali („bardzo niepewny” do „absolutnie pewny”) w jakim stopniu są przekonani o słuszności swojej oceny.

Analizy. Poziom zgodności oceniających kwantyfikowano z wykorzystaniem alfy Cronbacha po transpozycji macierzy — procedura zasugerowana przez McLennan (1993) pozwala łatwo ocenić zgodność pomiędzy licznymi oceniającymi. Aby określić trafność ocen, wynik oceny wiarygodności uzyskany przez każdego z oceniających porównywano z decyzją o prawdopodobieństwie podjętą niezależnie w czasie wyboru przesłuchań do badania.

Tabela 1. Rozkład ocen wiarygodności zależnie od wykorzystania protokołu i prawdopodobieństwa zdarzenia

Wykorzystanie protokołu	Oceny przesłuchujących					Łącznie
	Bardzo możliwe	Całkiem możliwe	Jednoznaczna ocena niemożliwa	Raczej niemożliwe	Absolutnie niemożliwe	
Przesłuchania nie-protokołowe						
Przypadki nieprawdopodobne	2 (4,8%)	15 (35,7%)	20 (47,6%)	5 (11,9%)	0 (0%)	42 (100%)
Przypadki prawdopodobne	7 (16,7%)	9 (21,4%)	24 (57,1%)	2 (4,8%)	0 (0%)	42 (100%)
Łącznie	9 (10,7%)	24 (28,6%)	44 (52,4%)	7 (8,3%)	0 (0%)	84 (100%)
Przesłuchania protokołowe						
Przypadki nieprawdopodobne	8 (19,0%)	12 (28,6%)	12 (28,6%)	10 (23,8%)	0 (0%)	42 (100%)
Przypadki prawdopodobne	28 (66,7%)	12 (28,6%)	2 (4,8%)	0 (0%)	0 (0%)	42 (100%)
Łącznie	36 (42,9%)	24 (28,6%)	14 (16,7%)	10 (11,9%)	0 (0%)	84 (100%)

Oceny wiarygodności poddano następnie analizie w odniesieniu do prawdopodobieństwa danego zdarzenia i zastosowanej formy przesłuchania (według protokołu NICHD czy bez protokołu). W analizach wykorzystano testy nieparametryczne Friedmanna lub McNemara (dla zmiennych dichotomicznych) dla prób zależnych (Hays, 1973). Porównanie poziomu pewności prze-

słuchujących w odniesieniu do formułowanych opinii w zależności od dwóch czynników (wykorzystanie protokołu i prawdopodobieństwo zdarzenia) przeprowadzono przy użyciu analizy wariancji (ANOVA) z powtarzanymi pomiarami. Zależność pomiędzy trafnością i pewnością oceny określano z wykorzystaniem współczynników korelacji Pearsona.

Wyniki

Rzetelność oceniana metodą zgodności pomiędzy oceniającymi

Rozkład ocen wiarygodności, jako funkcję wykorzystania protokołu i prawdopodobieństwa zdarzeń, przedstawiono w tabeli 1. Zaprezentowane dane wskazują, że opcja „zdarzenie absolutnie niemożliwe” nie została ani razu wybrana, ani w ocenie przesłuchań zgodnych z protokołem, ani w ocenie przesłuchań nie-protokołowych. W ocenie przesłuchań nieuwzględniających protokołu rozkład opinii był większy, a poziom zgodności między oceniającymi niższy ($\alpha = 0,764$) niż w przypadku przesłuchań we-

ługprotokołu ($\alpha = 0,874$). Większa rzetelność ocen w przypadku przesłuchań protokołowych niż nie-protokołowych była szczególnie widoczna w ocenach zarzutów nieprawdopodobnych ($\alpha = 0,642$ dla protokołowych i $\alpha = 0,338$ dla nie-protokołowych) i mniej wyraźna w ocenie zarzutów prawdopodobnych ($\alpha = 0,811$ dla protokołowych i $\alpha = 0,890$ dla nie-protokołowych). Analiza χ^2 potwierdziła, że rozkład ocen przesłuchań według protokołu i przesłuchań nieprotokołowych był różny ($p < 0,0001$).

Wykorzystanie opcji „jednoznaczna ocena niemożliwa” (JON)

Jak wykazano w tabeli 2, prawie jedna szóstka (16,7%) ocen odnoszących się do przesłuchań według protokołu uwzględniała wykorzystanie kategorii JON — zastosowano ją w 4,8% przypadków prze-

słuchań według protokołu dotyczących zarzutów prawdopodobnych i w 28,6% przypadków przesłuchań według protokołu dotyczących zarzutów nieprawdopodobnych.

Tabela 2. Częstość stosowania kategorii „jednoznaczna ocena niemożliwa” (JON) zależnie od wykorzystania protokołu i prawdopodobieństwa zdarzenia.

Wykorzystanie protokołu	Wykorzystanie opcji JON		Łącznie
	Ocena możliwa	Ocena niemożliwa	
Przesłuchania nie-protokołowe			
Przypadki nieprawdopodobne	22 (52,4%)	20 (47,6%)	42 (100%)
Przypadki prawdopodobne	18 (42,9%)	24 (57,1%)	42 (100%)
Łącznie	40 (47,6%)	44 (52,4%)	84 (100%)
Przesłuchania protokołowe			
Przypadki nieprawdopodobne	30 (71,4%)	12 (28,6%)	42 (100%)
Przypadki prawdopodobne	40 (95,2%)	2 (4,8%)	42 (100%)
Łącznie	70 (83,3%)	14 (16,7%)	84 (100%)

Dla porównania, ponad połowa (52,4%) ocen przesłuchań nieuwzględniających protokołu — w tym 57,1% ocen w przypadkach prawdopodobnych i 47,6% ocen w przypadkach nieprawdopodobnych — wykorzystywała kategorię JON. Test nieparametryczny McNemara dla prób zależnych potwierdził,

że kategoria JON pojawiała się częściej w ocenach przesłuchań nie-protokołowych niż protokołowych ($p < 0,019$). Efekt ten był wyraźny (bliski istotności, $p < 0,06$), gdy do analizy włączono tylko przypadki prawdopodobne, i mniej wyraźny, gdy analiza dotyczyła wyłącznie przypadków nieprawdopodobnych.

Trafność ocen

Prawie 60% (59,5%) ocen przesłuchań prowadzonych z wykorzystaniem protokołu — w tym 95,2% ocen w przypadkach prawdopodobnych i 23,8% ocen w przypadkach nieprawdopodobnych — okazało się trafnych (tzn. oceniający właściwie ocenił relację jako wiarygodną lub niewiarygodną); patrz tabela 3. Dla porównania, mniej niż jedna trzecia (29,6%) ocen odnoszących się do przesłu-

chań prowadzonych bez protokołu — w tym 38,1% ocen dotyczących przypadków prawdopodobnych i 11,9% ocen w przypadkach nieprawdopodobnych — była trafna. Zarówno zarzuty prawdopodobne ($p < 0,0001$), jak i nieprawdopodobne ($p < 0,0001$) były częściej oceniane trafnie, jeśli ocena dotyczyła przesłuchań prowadzonych według protokołu niż bez niego.

Tabela 3. Częstość trafnych i nietrafnych ocen wiarygodności zależnie od wykorzystania protokołu i prawdopodobieństwa zdarzenia (w tym JON).

Wykorzystanie protokołu	Trafność oceny		Łącznie
	Nietrafna	Trafna	
Przesłuchania nie-protokołowe			
Przypadki nieprawdopodobne	37 (88,1%)	5 (11,9%)	42 (100%)
Przypadki prawdopodobne	26 (61,9%)	16 (38,1%)	42 (100%)
Łącznie	63 (75,0%)	21 (25,0%)	84 (100%)
Przesłuchania protokołowe			
Przypadki nieprawdopodobne	32 (76,2%)	10 (23,8%)	42 (100%)
Przypadki prawdopodobne	2 (4,8%)	40 (95,2%)	42 (100%)
Łącznie	34 (40,5%)	50 (59,5%)	84 (100%)

Tabela 4. Poziom pewności oceniających odnośnie do własnych ocen wiarygodności zależnie od wykorzystania protokołu i prawdopodobieństwa zdarzenia.

Wykorzystanie protokołu	Średnia	SD	N
Przesłuchania nie-protokołowe Prawdopodobieństwo			
Przypadki Nieprawdopodobne	3,52	0,97	42
Przypadki Prawdopodobne	3,93	0,78	42
Łącznie	3,72	0,89	84
Przesłuchania protokołowe Prawdopodobieństwo			
Przypadki Nieprawdopodobne	3,90	0,58	42
Przypadki Prawdopodobne	4,19	0,55	42
Łącznie	4,04	0,57	84

Poziom pewności oceniających odnośnie do własnej oceny

W wyniku analizy wariancji (ANOVA) z powtarzonymi pomiarami ustalono, że zarówno zastosowanie protokołu, jak i dodatnie prawdopodobieństwo danego przypadku w istotny sposób wpływały na poziom pewności, jaką osoby oceniające pokładały we własnej ocenie wiarygodności (dla wpływu protokołu $F_{(1,41)}=11,90$; $p=0,001$; dla wpływu prawdopodobieństwa $F_{(1,41)}=6,61$; $p=0,014$).

Ocena przesłuchań przebiegających zgodnie z protokołem wiązała się z wyższym poziomem pewności niż ocena przesłuchań, w których protokołu nie uwzględniono. Analogicznie, zarzuty prawdopodobne oceniano z większą dozą pewności niż zarzuty nieprawdopodobne (patrz tabela 4). Nie zauważono istotnej interakcji pomiędzy tymi dwoma czynnikami.

Zależność trafność–pewność

Współczynnik korelacji Pearsona pomiędzy deklarowanym stopniem pewności dotyczącej oceny a faktyczną trafnością oceny wiarygodności, w przypadku przesłuchań przeprowadzonych bez uwzględnienia protokołu nie wykazał istotnych powiązań pomiędzy trafnością a pewnością, niezależnie od poziomu prawdopodobieństwa relacji ($r = -0,07$ w przypadku relacji prawdopodob-

nych oraz $r = 0,07$ w przypadku relacji nieprawdopodobnych). W odniesieniu do przesłuchań przebiegających zgodnie z protokołem, zaobserwowano natomiast dodatnią i istotną korelację pomiędzy trafnością oceny a poziomem pewności w przypadku zeznań prawdopodobnych ($r = 0,37$; $p < 0,015$). W przypadku zeznań nieprawdopodobnych natomiast zależność ta była nieistotna ($r = -0,12$).

Dyskusja

Celem niniejszego badania było rozstrzygnięcie, czy zastosowanie protokołu dochodzeniowego NICHHD w trakcie przesłuchania może poprawić trafność oceny wiarygodności i poziom zgodności między niezależnymi oceniającymi (*inter-rater reliability*), w porównaniu do przesłuchań odbywających się bez uwzględnienia protokołu. Wyniki wykazały, że w toku przesłuchań przebiegających w zgodzie z protokołem doświadczeni specjaliści dochodzeniowi dwukrotnie częściej niż w przypadku przesłuchań o odmiennej strukturze trafnie oceniali wiarygodność zeznań dzieci (59,5% vs. 29,6%). Tak wyraźny oddźwięk zastosowania protokołu nabiera szczególnej wagi w obliczu faktu, że żaden ze stosowanych wcześniej środków nie wykazywał podobnego wpływu na ocenę wiarygodności, przez co specjaliści niejednokrotnie nie potrafili odróżnić zdarzeń faktycznie doświadczonych od tych, które nie miały miejsca (Ceci i wsp. 1994; Horner i wsp. 1993a, 1993b; Leichtman, Ceci 1995).

Protokół NICHHD zdaje się dodatnio wpływać na trafność oceny wiarygodności, ponieważ jakość zeznań dzieci jest wyższa, gdy przesłuchujący stosują się do zaleceń „dobrej praktyki”. Wcześniejsze badania, których celem była ocena protokołu NICHHD wykazały, że w toku przesłuchań przeprowadzanych w oparciu o protokół częściej wykorzystywano pytania otwarte, stymulujące swobodne odtwarzanie i pozwalające uzyskać bardziej szczegółowe relacje z doświadczenia krzywdy (Orbach i wsp. 2000; Sternberg, Lamb, Davies i wsp. 2001; Sternberg, Lamb, Orbach i wsp. 2001). Dzięki strategii pytań otwartych potrafimy zaobserwować więcej prawdopodobnych wskaźników wiarygodności. Dla przykładu, prawdopodobieństwo rozpoznania kryteriów CBCA w zeznaniach dzieci jest wyższe, jeżeli przesłuchujący stosuje wobec domniemych ofiar strategię pytań otwartych w celu uzy-

skania wypowiedzi narracyjnej (Hershkowitz i wsp. 1997). Ponadto, w porównaniu z pytaniami szczegółowymi, strategie otwarte pozwalają uzyskać bogatsze opisy zdarzeń w przypadku dzieci stawiających zarzuty prawdopodobne niż w przypadku dzieci formułujących oskarżenia nieprawdopodobne (Hershkowitz 1999).

Wyniki powyższego badania wskazują, że zastosowanie protokołu NICHHD nie tylko poprawia jakość przesłuchania, ale również ułatwia ocenę wiarygodności zeznań uzyskanych od dziecka. Choć protokół miał w zamyśle ułatwiać praktyczne zastosowanie wytycznych dobrej praktyki umożliwiających dzieciom składanie wysokiej jakości zeznań, jego szerszym przeznaczeniem była pomoc specjalistom dochodzeniowym w ewaluacji zeznań składanych przez dzieci, a tym samym poprawa sytuacji dzieci krzywdzonych w kontekście wymiaru sprawiedliwości. Prezentowane wyniki dowodzą, że protokół NICHHD jest pod tym względem cennym narzędziem praktyki sądowej.

Na szczególną uwagę zasługuje również wysoki stopień rozbieżności pomiędzy trafnością ocen dotyczących zeznań prawdopodobnych a nieprawdopodobnych. Zeznania nieprawdopodobne były zdecydowanie rzadziej słusznie identyfikowane jako niewiarygodne niż zeznania prawdopodobne jako wiarygodne. Inni badacze (Akehurst i wsp. 2001; Steller i wsp. 1988; Tye i wsp. 1995; Yuille 1988) także wskazują, że zarówno opiniujący specjaliści, jak i studenci bardziej trafnie oceniają zeznania prawdziwe bądź prawdopodobne niż nieprawdopodobne czy fałszywe. Niemniej jednak zaskoczył nas fakt, że nawet osoby doświadczone w ocenie, rozpoznawały zeznania nieprawdopodobne z częstością graniczącą z przypadkiem.

Można by wysunąć argument, że czynnikiem istotnie wpływającym na niezadowalające wyniki specjalistów oceniających był

dostęp jedynie do słownego zapisu przesłuchania, bez nagrań wideo czy bezpośredniej obserwacji dzieci. Inni badacze zauważają, że w celu identyfikacji zeznań prawdziwych specjaliści w większym stopniu kierują się informacjami werbalnymi, komunikaty fałszywe są natomiast częściej rozpoznawane na podstawie sygnałów pozawerbalnych (Anderson, De Paulo, Ansfield 1999; Anderson, De Paulo, Ansfield i wsp. 1999; Vrij, Edward i wsp. 2001). Można by więc sugerować, że brak komunikatów pozawerbalnych w niniejszym badaniu w sposób wybiórczy osłabił zdolność oceniających specjalistów do klasyfikacji zeznań nieprawdopodobnych jako niewiarygodnych. To wytłumaczenie zdaje się mało przekonujące, ponieważ domniemanie szczerości („*truth bias*”) było wyraźnie obecne w licznych wcześniejszych badaniach, niezależnie od tego, czy wykorzystywano w nich komunikaty niewerbalne czy też nie (przegląd patrz: Vrij, 2000).

Wiele czynników ogranicza możliwość uogólnienia przedstawionych tu spostrzeżeń. Po pierwsze, istnieje możliwość, że w praktyce sądowej specjaliści dochodzeniowi, mając dostęp do sygnałów pozawerbalnych trafniej oceniają wiarygodność zeznań. Po drugie, choć wyniki uzyskiwane przez oceniających w rozpoznawaniu oskarżeń nieprawdopodobnych jako niewiarygodnych były niezadowolające, w rozpatrywanych przypadkach częściej stwierdzano niemożność wydania opinii niż błędnie zakładano wiarygodność oskarżenia. W rzeczywistym kontekście podobna ocena mogłaby sprowokować do dalszego poszukiwania informacji, czy to w formie ponownego przesłuchania dziecka, domagania się badań lekarskich czy próby zdobycia informacji od świadków. Powyższe kroki mogłyby ostatecznie dodatnio wpłynąć na prawdopodobieństwo trafnej oceny wiarygodności zarzutów. Po trzecie, w badaniu uwzględniono taką samą liczbę przypadków prawdopodobnych i nieprawdopodobnych, w rzeczy-

wistości natomiast ogromna większość zarzutów stawianych przez dzieci w kontekście dochodzeniowym zdaje się być prawdopodobna (London, Bruck, Ceci, Shuman 2005). Przeświadczenie, że większość zarzutów jest na ogół prawdopodobna mogła wpłynąć na niechęć specjalistów dochodzeniowych do określania zbyt wielu zapisów jako nieprawdopodobnych. Pomimo powyższych zastrzeżeń, niezdolność do identyfikacji zarzutów nieprawdopodobnych przez osoby prowadzące dochodzenie budzi niepokój. Jakkolwiek trafność z jaką rozpoznawano relacje prawdopodobne sugeruje, że dzieci krzywdzone cieszą się odpowiednią ochroną, nieumiejętność odróżnienia przypadków nieprawdopodobnych oznacza, że fałszywe zarzuty mogą nieodpowiednio ukierunkować działania na rzecz ochrony dziecka i postępowanie wymiaru sprawiedliwości, co w rezultacie zaowocuje poważnymi konsekwencjami.

Przesłuchania przeprowadzone w oparciu o protokół, dostarczające zarówno prawdopodobnych, jak i nieprawdopodobnych zarzutów, zostały ocenione trafniej niż przesłuchania przeprowadzone bez uwzględnienia protokołu. W odniesieniu do zarzutów prawdopodobnych, zastosowanie protokołu niemal kompletnie wyeliminowało przypadki nietrafnego opiniowania. Błędne oceny zdarzały się jednak nadal alarmująco często w odniesieniu do relacji nieprawdopodobnych. Żadne ze skrzywdzonych dzieci przesłuchiowanych w oparciu o protokół nie zostało błędnie oskarżone o składanie fałszywych zeznań. Natomiast dzieci, które nie padły ofiarą krzywdy, a składały fałszywe zeznania, niepokojąco często uznawane były za faktyczne ofiary przez tych samych doświadczonych specjalistów.

W przypadku przesłuchań zgodnych z protokołem, wysoki poziom zgodności między niezależnymi oceniającymi odzwierciedla wysoki poziom trafności ocen i wymaga szczególnego podkreślenia. Poziom

zgodności w ocenie przesłuchań opartych o protokół był zaskakująco wysoki, zarówno w porównaniu do zgodności w ocenie przesłuchań nieuwzględniających protokołu, jak i w porównaniu z analogicznymi wielkościami otrzymywanymi we wcześniejszych badaniach (Finlayson, Koocher 1991; Horner i wsp. 1993a, 1993b; Jackson, Nuttal 1993; Poole, Lindsay 1997). Wysoka zgodność nie jest wartością samą w sobie, gdyż oceniający mogą być zgodni również w odniesieniu do osądów niewłaściwych (co miało miejsce niejednokrotnie odnośnie do przesłuchań prowadzonych bez zastosowania protokołu), niemniej wysoka zgodność w odniesie-

niu do właściwych ocen zarzutów postawionych w trakcie przesłuchań opartych na protokole świadczy o dodatkowych zaletach tego narzędzia. Uzyskane w niniejszym badaniu wyniki wskazują w szczególności, że zastosowanie protokołu nie tylko korzystnie wpływa na jakość przesłuchania dochodzeniowego (Orbach i wsp. 2000; Sternberg, Lamb, Davies i wsp. 2001; Sternberg, Lamb, Orbach i wsp. 2001), ale niweluje też indywidualne różnice w procesie przesłuchiwania i ocenie wiarygodności, zwiększając tym samym szanse na pomyślne i sprawiedliwe zakończenie sprawy, niezależnie od tego, kto prowadzi przesłuchanie.

Podziękowania

Autorzy dziękują śledczym specjalizującym się w pracy z młodzieżą za ich udział w badaniu oraz Wydziałowi ds. Środków Poprawczych za nieustające wsparcie niniejszego programu badawczego.

Tłumaczenie: Anna Czajkowska-Błaziak

Przedruk za zgodą wydawcy, tłumaczenie redakcyjne, niekonsultowane z wydawcą. Artykuł opublikowano w: Child Abuse and Neglect, vol. 31, *Improving credibility assessment in child sexual abuse allegations: The role of the NICHD investigative interview protocol*, s. 99–110, Copyright Elsevier (2007).

Objectives. *The study was designed to explore whether the credibility of children's statements regarding their alleged experiences of child sexual abuse could be assessed in a more valid and reliable way when investigative interviews were conducted using the NICHD protocol rather than in an unstructured manner.*

Methods. *Forty-two experienced Israeli youth investigators each assessed the credibility of allegations of sexual abuse made by alleged victims of sexual abuse when interviewed either with or without the protocol. Half of the alleged incidents were judged likely to have happened ("plausible") on the basis of independent evidence, while half were deemed unlikely to have happened ("implausible").*

Results. *More non-protocol than protocol interviews were rated as "No judgment possible" rather than either credible or incredible. Allegations made in protocol interviews were more accurately rated as credible or incredible when they were either plausible or implausible, respectively, than those made in non-protocol statements. Levels of inter-rater reliability were also higher when protocol interviews were rated. The differences were significant only for plausible cases, however.*

Conclusions. *The use of the NICHD protocol facilitated the assessment of credibility by child investigators although incredible allegations (those describing incidents that were unlikely to have happened) remained difficult to detect, even when the protocol was used.*

Literatura

- Akehurst L., Koehnken G., Hofer E. (2001), *Content credibility of accounts derived from live and video presentations*, *Legal and Criminological Psychology*, 6, s. 65–83.
- Anderson D.E., De Paulo B.M., Ansfield M.E. (1999), *The development of deception detection skill: A longitudinal study of same sex friends*, niepublikowany rękopis, Uniwersytet w Wirginii, Charlottesville.
- Anderson D.E., De Paulo B.M., Ansfield M.E., Tickle J.J., Green E. (1999), *Beliefs about cues to deception: Mindless stereotypes or untapped wisdom?* *Journal of Nonverbal Behavior*, 23, s. 67–89.
- Boychuk T.D. (1991), *Criteria-based content analysis of children's statements about sexual abuse: A field-based validation study*, niepublikowana praca doktorska, Uniwersytet Stanowy w Arizonie, Tempe, AZ.
- Ceci S.J., Crotteau-Huffman M.L. (1997), *How suggestible are preschool children? Cognitive and social factors*, *Journal of the American Academy of Child and Adolescent Psychiatry*, 36, s. 948–958.
- Ceci S.J., Loftus E.F., Leichtman M.D., Bruck M. (1994), *The possible role of source misattributions in the creation of false beliefs among preschoolers*, *International Journal of Clinical and Experimental Hypnosis*, 42, s. 304–320.
- Cederborg A.C., Orbach Y., Sternberg K.J., Lamb M.E. (2000), *Investigative interviews of child witnesses in Sweden*, *Child Abuse & Neglect*, 24, s. 1355–1361.
- Chahal K., Cassidy T. (1995), *Deception and its detection in children: A study of adult accuracy, psychology*, *Crime & Law*, 1, s. 237–245.
- Craig R.A., Sheibe R., Kircher J., Raskin D.C., Dodd D. (1999), *Effects of interviewer questions on children's statements of sexual abuse*, *Applied Developmental Science*, 3, s. 77–85.
- DePaulo B.M., Charlton K., Cooper H., Lindsay J.L., Muhlenbruck L. (1997), *The accuracy–confidence correlation in the detection of deception*, *Personality and Social Psychology Review*, 1, s. 346–357.
- Finlayson L.M., Koocher G.P. (1991), *Professional judgment and child abuse reporting in sexual abuse cases*, *Professional Psychology: Research and Practice*, 22, s. 464–472.
- Hays W.L. (1973), *Statistics for the social sciences* (wyd. 2), Nowy Jork: Holt, Rinehart i Winston.
- Hershkowitz I. (1999), *The dynamics of interviews involving plausible and implausible allegations of sexual abuse*, *Applied Developmental Science*, 3, s. 86–91.
- Hershkowitz I., Lamb M.E., Sternberg K.J., Esplin P.W. (1997), *The relationships among interviewer utterance type, CBCA scores, and the richness of children's responses*, *Legal and Criminological Psychology*, 2, s. 169–176.
- Horner T.M., Guyer M.J., Kalter N.M. (1993a), *Clinical expertise and the assessment of child sexual abuse*, *Journal of the American Academy of Child and Adolescent Psychiatry*, 32, s. 925–931.
- Horner T.M., Guyer M.J., Kalter N.M. (1993b), *The biases of child sexual abuse experts: Believing is seeing*, *Bulletin of the American Academy of Psychiatry and the Law*, 21, s. 281–292.
- Horowitz S.W., Lamb M.E., Esplin P.W., Boychuk T.D., Krispin O., Reiter-Lavery L. (1997), *Reliability of criteria-based content analysis of child witness statements*, *Legal and Criminological Psychology*, 2, s. 11–21.
- Horowitz S.W., Lamb M.E., Esplin P.W., Reiter-Lavery L., Krispin C. (1995), *Establishing ground truth in studies of child sexual abuse*, *Expert Evidence*, 4, s. 42–51.

- Jackson H., Nuttall R. (1993), *Clinician responses to sexual abuse allegations*, *Child Abuse & Neglect*, 17, s. 127–143.
- Jackson J.L. (1996), *Truth or fantasy: The ability of barristers and laypersons to detect deception in children's testimony*, odczyt wygłoszony na konferencji Amerykańskiego Towarzystwa Psychologii i Prawa (AP-LS), Hilton Head Island, Karolina Południowa.
- Lamb M.E., Hershkowitz I., Sternberg K.J., Esplin P.W., Hovav M., Manor T., Yudilevitch L. (1996), *Effects of investigative style on Israeli children's responses*, *International Journal of Behavioral Development*, 19, s. 627–637.
- Lamb M.E., Sternberg K.J., Esplin P.W., Hershkowitz I., Orbach Y., Hovav M. (1997), *Criterion-based content analysis: A field validation study*, *Child Abuse & Neglect*, 21, s. 255–264.
- Lamb M.E., Sternberg K.J., Orbach Y., Esplin P.W., Stewart H., Mitchell S. (2003), *Age differences in young children's responses to open-ended invitations on the course of forensic interviews*, *Journal of Consulting and Clinical Psychology*, 71, s. 926–934.
- Leichtman M.D., Ceci S.J. (1995), *The effects of stereotypes and suggestion on preschoolers' reports*, *Developmental Psychology*, 31, s. 568–578.
- Lewis M. (1993), *The development of deception*, w: M. Lewis, C. Saarni (red.), *Lying and deception in everyday life* (s. 90–105), Nowy Jork: Guilford Press.
- Lewis M., Stanger C., Sullivan M.W. (1989), *Deception in 3-year-olds*, *Developmental Psychology*, 25, s. 439–443.
- London K., Bruck M., Ceci S.J., Shuman D.W. (2005), *Disclosure of child sexual abuse: What does the research tell us about the ways that children tell?* *Psychology, Public Policy, and the Law*, 11, s. 194–226.
- McLennan R.N. (1993), *Inter-rater reliability with SPSS for Windows 5.0.*, *The American Statistician*, 47(4), s. 292–297.
- Orbach Y., Hershkowitz I., Lamb M.E., Sternberg K.J., Esplin P.W., Horowitz D. (2000), *Assessing the value of structured protocols for forensic interviews of alleged child abuse victims*, *Child Abuse & Neglect*, 24, s. 733–752.
- Poole D.A., Lindsay D.S. (1997, kwiecień), *Misinformation from parents and children's source monitoring: Implications for testimony*, odczyt wygłoszony dla Stowarzyszenia na rzecz Badań nad Rozwojem Dziecka (Society for Research in Child Development), Waszyngton, DC.
- Raskin D.C., Esplin P.W. (1991a), *Assessment of children's statements of sexual abuse*, w: J. Doris (red.), *The Suggestibility of children's recollections* (s. 153–164), Waszyngton, DC: Amerykańskie Towarzystwo Psychologiczne.
- Raskin D.C., Esplin P.W. (1991b), *Statement Validity Assessment: Interview procedures and content analysis of children's statements of sexual abuse*, *Behavioral Assessment*, 12, s. 265–291.
- Steller M., Wellerhaus P., Wolf T. (czerwiec, 1988), *Empirical validation of criteria based content analysis*, odczyt wygłoszony w Instytucie Badań nad Oceną Wiarygodności przy NATO (NATO Advanced Study Institute on Credibility Assessment), Maratea, Włochy.
- Sternberg K.J., Lamb M.E., Davies G.M., Westcott H.L. (2001), *The Memorandum of Good Practice: Theory versus application*, *Child Abuse & Neglect*, 25, s. 669–681.
- Sternberg K.J., Lamb M.E., Hershkowitz I. (1996), *Child sexual abuse investigations in Israel*, *Criminal Justice & Behavior*, 23, s. 322–337.
- Sternberg K.J., Lamb M.E., Orbach Y., Esplin P.W., Mitchell S. (2001), *Use of a structured investigative protocol enhances young children's responses to free recall prompts in the course of forensic interviews*, *Journal of Applied Psychology*, 85, s. 997–1005.

- Tye M.J.C., Henderson S.A., Honts C.R. (1995, styczeń), *Evaluating children's testimonies: CBCA and lay subjects*, odczyt wygłoszony na spotkaniu CRIME CON: Międzynarodowej Konferencji Internetowej ds. Przestępczości i Wymiaru Sprawiedliwości w Sprawach Karnych (International Internet Conference on Crime and Criminal Justice), Little Rock, AR.
- Vrij A. (2000), *Detecting lies and deceit: The psychology of lying and the implications for Professional practice*, Wiley: Chichester, Wielka Brytania.
- Vrij A., Baxter M. (2000), *Accuracy and confidence in detecting truths and lies in elaborations and denials: Truth bias, lie bias and individual differences*, *Expert Evidence*, 7, s. 25–36.
- Vrij A., Edward K., Bull R. (2001), *Police officers ability to detect deceit: the benefit of indirect deception detection measures*, *Legal and Criminological Psychology*, 6, s. 185–196.
- Vrij A., Harden F., Terry J., Edward K., Bull R. (2001), *The influence of personal characteristics, stakes and lie complexity on the accuracy and confidence to detect deceit*, w: R. Roesch, R.R. Corrado, R.J. Dempster (red.), *Psychology in the courts: International advances in knowledge*, Amsterdam: Harwood Academic.
- Vrij A., Van Wijngaarden J.J. (1994), *Will truth come out? Two studies about the detection of false statements expressed by children*. *Expert evidence: The International digest of Human Behavior, Science and Law*, 3, s. 78–84.
- Wells G.L., Loftus E.F. (1991), *Commentary: Is this child fabricating? Reactions to a new assessment technique*, w: J. Doris (red.), *The suggestibility of children's recollections* (s. 168–171), Waszington, DC: Amerykańskie Towarzystwo Psychologiczne.
- Westcott H.L., Davies G.M., Clifford B.R. (1991), *Adults' perceptions of children's videotaped truthful and deceptive statements*, *Children & Society*, 5, s. 123–135.
- Yuille J.C. (1988), *The systematic assessment of children's testimony*, *Canadian Psychologist*, 29, s. 247–262.

O AUTORACH

MICHAEL E. LAMB przewodził Sekcji ds. Rozwoju Społecznego i Emocjonalnego w amerykańskim Narodowym Instytucie Zdrowia Dziecka i Rozwoju Człowieka (National Institute of Child Health and Human Development), zanim został profesorem psychologii na Uniwersytecie w Cambridge. Doktorat uzyskał na Uniwersytecie Yale, otrzymał doktoraty *honoris causa* uniwersytetów w Goteborgu i East Anglia. Jest autorem lub redaktorem ponad 40 książek, a także około 500 publikacji naukowych.

IRIT HERSHKOWITZ jest starszym wykładowcą i pracownikiem naukowym w Szkole Pracy Społecznej na Uniwersytecie w Hajfie w Izraelu, na której uzyskała doktorat przed zaakceptowaniem podoktorskiego stypendium w amerykańskim Narodowym Instytucie Zdrowia Dziecka i Rozwoju Człowieka (National Institute of Child Health and Human Development). Jej badania koncentrowały się na reakcjach dzieci dotyczących różnych strategii dochodzeniowych, na wartości szkolenia detektywów dziecięcych, na dyskryminacji podstawnych i bezpodstawnych zarzutów wnoszonych przez dzieci oraz na rozwoju techniki przesłuchiwania świadków nieskłonnych do współpracy i osób szczególnej troski.