What Works for Children's Social Care

CARDIFF UNIVERSITY PRIFYSGOL CAERDYDD

A RANDOMISED CONTROLLED TRIAL OF A 'CHECKLIST' INTERVENTION TO MITIGATE CONFIRMATION BIAS AND IMPROVE FORECASTING ACCURACY IN SOCIAL WORK

August 2021

What Works *for* Children's Social Care

Protocol registration

The full trial protocol was registered with Open Science Framework (osf.io) on 4th December 2020 (https://osf.io/57b84).

Funding

Department for Education, England.

Authors

Wilkins, D., CASCADE, School of Social Sciences, Cardiff University; **Meindl, M.**, CASCADE, School of Social Sciences, Cardiff University

About What Works for Children's Social Care

What Works for Children's Social Care seeks better outcomes for children, young people and families by bringing the best available evidence to practitioners and other decision makers across the children's social care sector. We generate, collate and make accessible the best evidence for practitioners, policy makers and practice leaders to improve children's social care and the outcomes it generates for children and families.

About CASCADE

CASCADE is concerned with all aspects of community responses to social need in children and families, including family support services, children in need services, child protection, looked after children and adoption. It is the only centre of its kind in Wales and has strong links with policy and practice.

Visit WWCSC at: whatworks-csc.org.uk or
CASCADE at: sites.cardiff.ac.uk/cascade

If you'd like this publication in an alternative format such as Braille, large print or audio, please contact us at: info@whatworks-csc.org.uk

CONTENTS

EXECUTIVE SUMMARY

Social workers make potentially life-changing decisions every day – including whether to accept a referral, undertake a child protection investigation, pursue care proceedings or close the case. Many of these decisions involve implicit or explicit predictions about the likelihood of different future outcomes. If the case is closed, will the child and family remain safe and well? If we pursue care proceedings, will this help protect the child from harm, or could the same outcome be achieved via a voluntary agreement with the child's parents? If the family are referred to a particular service, will they find it helpful, and to what extent? In addition, the law in England (and Wales) requires that social workers make judgements about the likelihood of future significant harm. Such judgements involve making "*a prediction from existing facts, often from a multitude of such facts, about what has happened in the past, about the characters and personalities of the people involved,* [and] *about the things which they have said and done*" (Parliament. House of Lords, 2008).

Social workers make potentially life-changing decisions every day – including whether to accept a referral, undertake a child protection investigation, pursue care proceedings or close the case. Many of these decisions involve implicit or explicit predictions about the likelihood of different future outcomes. If the case is closed, will the child and family remain safe and well? If we pursue care proceedings, will this help protect the child from harm, or could the same outcome be achieved via a voluntary agreement with the child's parents? If the family are referred to a particular service, will they find it helpful, and to what extent? In addition, the law in England (and Wales) requires that social workers make judgements about the likelihood of future significant harm. Such judgements involve making "a prediction from existing facts, often from a multitude of such facts, about what has happened in the past, about the characters and personalities of the people involved, [and] about the things which they have said and done" (Parliament. House of Lords, 2008).

Yet predicting the future is not easy, and in studies from other fields, even recognised experts found it hard on average to perform better than you would expect by chance – and even harder to outperform a simple algorithm which predicted a continuation of the status quo (Tetlock, 2017; Tetlock & Gardner, 2016). One reason why people find it hard to make accurate forecasts is because of our susceptibility to cognitive bias. For example, we find it easier to recall times when our forecasts have proven accurate, and harder to remember when we have been wrong. This might result in an unjustified level of confidence in our own abilities. Similarly, we may find it easier to identify information in support

of our existing view, and harder to notice contradictory information. In social work, there are currently no well-evidenced interventions to help mitigate the effects of cognitive bias, although various effective interventions have been developed in other fields (Featherston et al., 2019).

In this study, we tested one intervention, a Checklist, and measured what difference it made in relation to i) forecasting abilities (measured using Brier scores) and ii) confirmation bias (measured using the Wason Selection Task).

Study design and sample

Eighty-eight participants (87 social workers and one student social worker) took part in an online survey, in which they were asked to read two baseline case studies and answer four questions about each one, to forecast the likelihood of different outcomes. Participants were then randomly allocated to a control group or an intervention group. Participants in the intervention group were asked to complete a checklist intervention, before answering questions about a further two endline case studies and completing the Wason Selection Task. All of the case studies were based on real-life referrals to social services, made approximately one year earlier, allowing us to know the actual outcomes of the forecasting questions and to assess the accuracy of participants' forecasts. For example, we asked participants to judge how likely it was for parents to attend appointments or be at home for visits, for the police to attend the family home, for professionals to report further concerns about the child and about where the child would be living and what type of social work plan they would be subject to over set periods of time following the referral.

The checklist intervention involved reading a longer vignette (not based on a real-life case) and making three recommendations about next steps. After reading the vignette and making an initial set of recommendations, participants were asked to work through an 8-item checklist. The items in the checklist aimed to help participants think about:

- How confident they felt about their recommendations
- If they had overlooked any disconfirming information in the case study, and
- Whether they would make the same decision if the family had different social or ethnic characteristics.

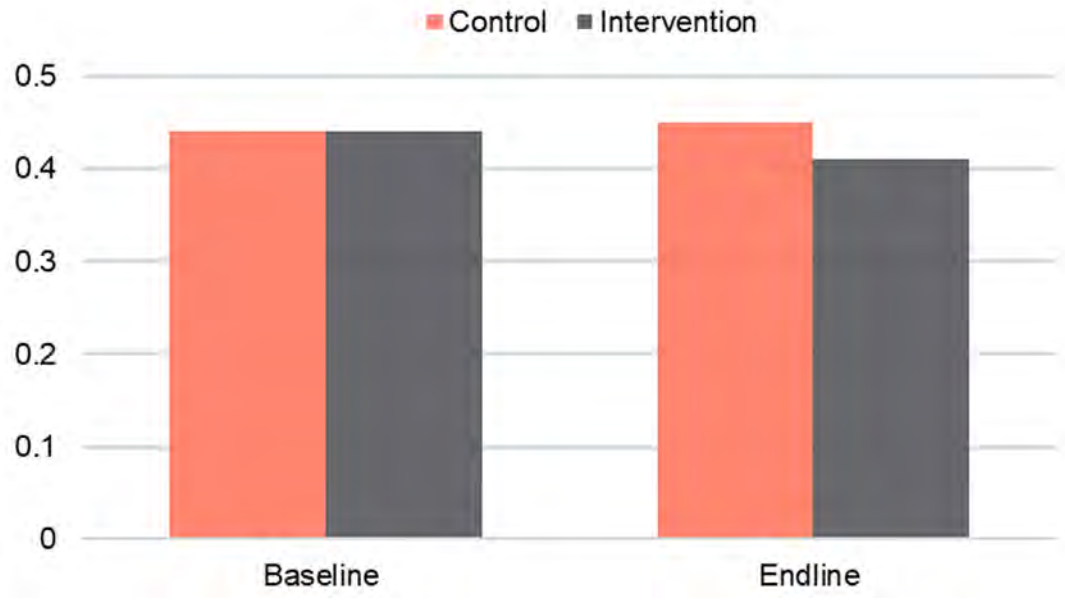
Participants were then given the opportunity to revise their recommendations, if they wanted to. Participants in the control group were asked to complete the two endline case studies and the Wason Selection Task before accessing the checklist intervention.

Findings

There was a post-intervention reduction in Brier scores for participants in the intervention group (indicating improved forecasting accuracy). However, this difference between the intervention and control groups was small and not statistically significant (*beta* = -.127, p = .23).

Comparison of Baseline and Endline Mean Brier scores

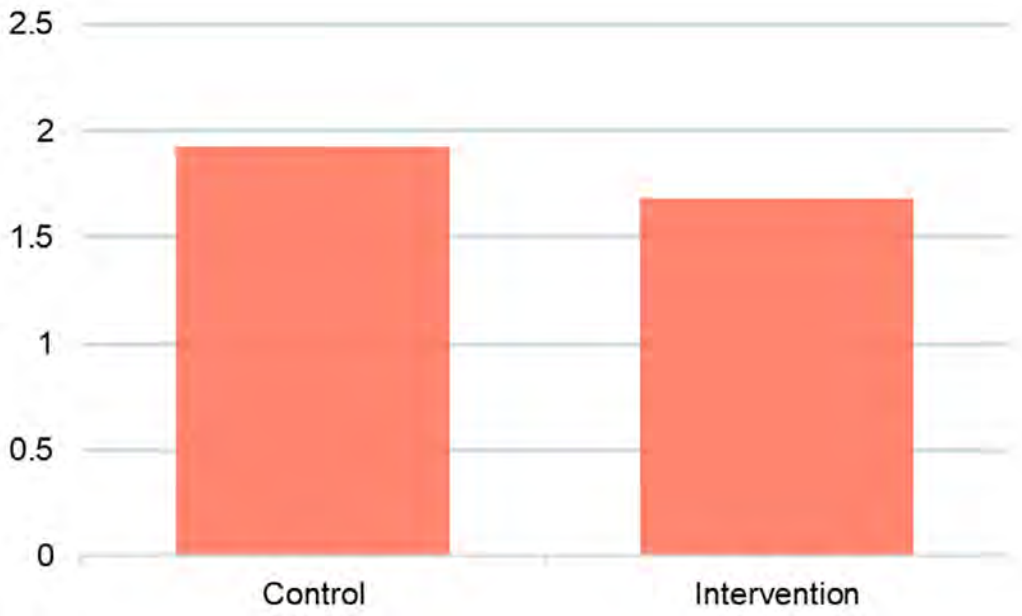
Participants' forecasting accuracy improved after the checklist intervention, but the change was not statistically significant



Trial of checklist intervention, n = 88

Between-group comparison of Wason Selection Task scores

The checklist intervention appeared to make no meaningful difference to confirmation bias



Trial of checklist intervention, n = 88

The checklist intervention also appeared to make no meaningful difference in relation to confirmation bias (*beta* = -.10, p = .38), with similar scores obtained from the Wason Selection Task between the two groups. The personal and professional characteristics of the participants were not associated with forecasting accuracy or confirmation bias.

Implications

Overall, we found no evidence that the checklist intervention made a significant difference in relation to forecasting accuracy or confirmation bias. These results, especially when considered alongside those of a previous study (Wilkins et al., 2020), suggest that interventions to improve forecasting accuracy (or mitigate cognitive bias) in social work would need to be much more in-depth than the relatively brief online interventions we have tested here and previously.

INTRODUCTION

Social workers make potentially life-changing decisions every day – including whether to accept a referral, undertake a child protection investigation, pursue care proceedings or close the case. Many of these decisions involve implicit or explicit predictions about the likelihood of different future outcomes. If the case is closed, will the child and family remain safe and well? If we pursue care proceedings, will this help protect the child from harm, or could the same outcome be achieved via a voluntary agreement with the child's parents? If the family are referred to a particular service, will they find it helpful, and to what extent? In addition, the law in England (and Wales) requires that social workers make judgements about the likelihood of future significant harm. Such judgements involve making "*a prediction from existing facts, often from a multitude of such facts, about what has happened in the past, about the characters and personalities of the people involved,* [and] *about the things which they have said and done*" (Parliament. House of Lords, 2008).

Yet predicting the future is not easy, and in studies from other fields, even recognised experts found it hard on average to perform better than you would expect by chance – and even harder to outperform a simple algorithm which predicted a continuation of the status quo (Tetlock, 2017; Tetlock & Gardner, 2016). One reason why people find it hard to make accurate forecasts is because of our susceptibility to cognitive bias (systematic errors in thinking that typically occur when we interpret information). For example, we find it easier to recall times when our forecasts have proven accurate, and harder to remember when we have been wrong. This might result in an unjustified level of confidence in our own abilities. Similarly, we may find it easier to identify information in support of our existing view, and harder to notice contradictory information. In social work, there are currently no well-evidenced interventions to help mitigate the effects of cognitive bias (Featherston et al., 2019).

Nevertheless, there are some reasons to be optimistic. In other fields, such as politics and economics, interventions have been developed which can significantly improve forecasting abilities (Tetlock, 2017; Tetlock & Gardner, 2016). According to the What Works for Children's Social Care website, a brief cognitive debiasing intervention helped to increase forecasting accuracy a modest amount – the equivalent of moving from the score of the 25th best forecaster out of 100, to the 10th best forecaster (14% of a standard deviation). This finding suggested that quick, low-cost and easy to administer interventions may help social workers avoid some of the negative effects of confirmation bias and improve forecasting accuracy.

In this study, we sought to explore this further by testing a slightly more in-depth intervention and measuring what difference it made in relation to i) forecasting abilities and ii) a direct measure of confirmation bias.

The intervention

In this study, we tested a *Checklist* intervention. Checklists have been widely used in other settings, such as healthcare, to improve decision-making (Gawande, 2010). In simple terms, a checklist consists of a series of prompts, and there is no one type of checklist suitable for universal application. Different examples are developed in a bespoke way, depending on the setting, the people involved, and the types of decisions being made. There is also no set format for what a checklist should include, although there are some broad principles, for example - the checklist should be as short as possible, employ simple language, and seek to prompt, rather than replace, professional judgement (Gawande, 2010).

The aim of any checklist is to ensure that the most important components of the decision-making process are not overlooked. Despite appearances, it is important to note that checklists are not necessarily a simple tool, and while they have helped to improve outcomes in relation to some forms of decision-making, they are far from being a panacea (Catchpole & Russ, 2015).

We intended to test the checklist intervention in-person; however due to the Covid-19 pandemic, we designed it to be delivered online. To our knowledge, checklists have not been implemented in this way before in social work. We thus designed a checklist and piloted it with a small group of social workers and social work academics (n=9). Following this pilot, we made some minor changes to the checklist and finalised the items (Figure 1).

Within the study itself, using the checklist involved reading a relatively long vignette and making three recommendations, two regarding what *should happen* next, and one about what *should not happen* (see Table 1 and Figure 1). Participants were then asked to read through the checklist to help them reflect on their initial recommendations and given the opportunity to revise them, if they wanted to.

Table 1: An overview of the vignette used as part of the Checklist intervention

Two White British children are referred – Sophie, aged 9 and Madeline, aged 3. Madeline was brought to hospital by her parents with burns to her neck, back and arms. The father said it was an accident but was also overheard telling the children not to talk to anyone about what really happened. You, the social worker, have been allocated to complete an assessment. When you visit the home, you find the family are struggling financially, and observe that the mother and children seem wary of the father. You also discover that the father has another child from a previous relationship, who lives in another part of the country with her mother. You check with the relevant local authority and find that this child used to live with her father and had a history of apparently accidental injuries. The next day, you come into work to find a report saying that police were called to the home last night because of concerns about domestic violence.

Concluding your assessment, what **two** recommendations would you make about what *should happen* now? What **one** thing would you recommend *should not happen*?

	Item	Yes	No	Unsure
1.	Have you identified two different options that you would recommend?			
2.	Have you identified one option you would not recommend?			
3.	Can you identify at least one pro and one con for all three recommendations?			
4.	Have you actively sought out information in the case study that might contradict your recommendations?			
5.	Can you see how another social worker might come up with a different set of recommendations?			
6.	Is it possible that your analysis of the case study is being influenced by cognitive biases such as confirmation bias?			
7.	Is it possible that you feel too confident about your recommendations?			
8.	Would you recommend the same options if the family were of a different ethnicity or social class?			

Figure 1: The checklist intervention used in this study.

Research Questions

Primary Research Questions

1. What is the impact of the checklist intervention on forecasting accuracy amongst social workers?
2. What is the impact of the checklist intervention on confirmation bias amongst social workers?

Exploratory Research Questions

3. Is there a relationship between social workers' forecasting accuracy and their level of confirmation bias?
4. Is there a relationship between social workers' forecasting accuracy and i) age-group, ii) gender, and iii) length of post-qualifying experience?

METHODS

The study was an individually randomised controlled trial with two arms, one intervention group and one control group, and two outcome measures, one pre and post, and one cross-over (Figure 2).

The *primary outcome* was the accuracy of forecasts measured by Brier scores, and the secondary outcome was the extent of confirmation bias measured by the Wason Selection Task.

The *primary hypothesis* is that endline Brier scores in the intervention group will be lower than in the control group, indicating more accurate forecasts.

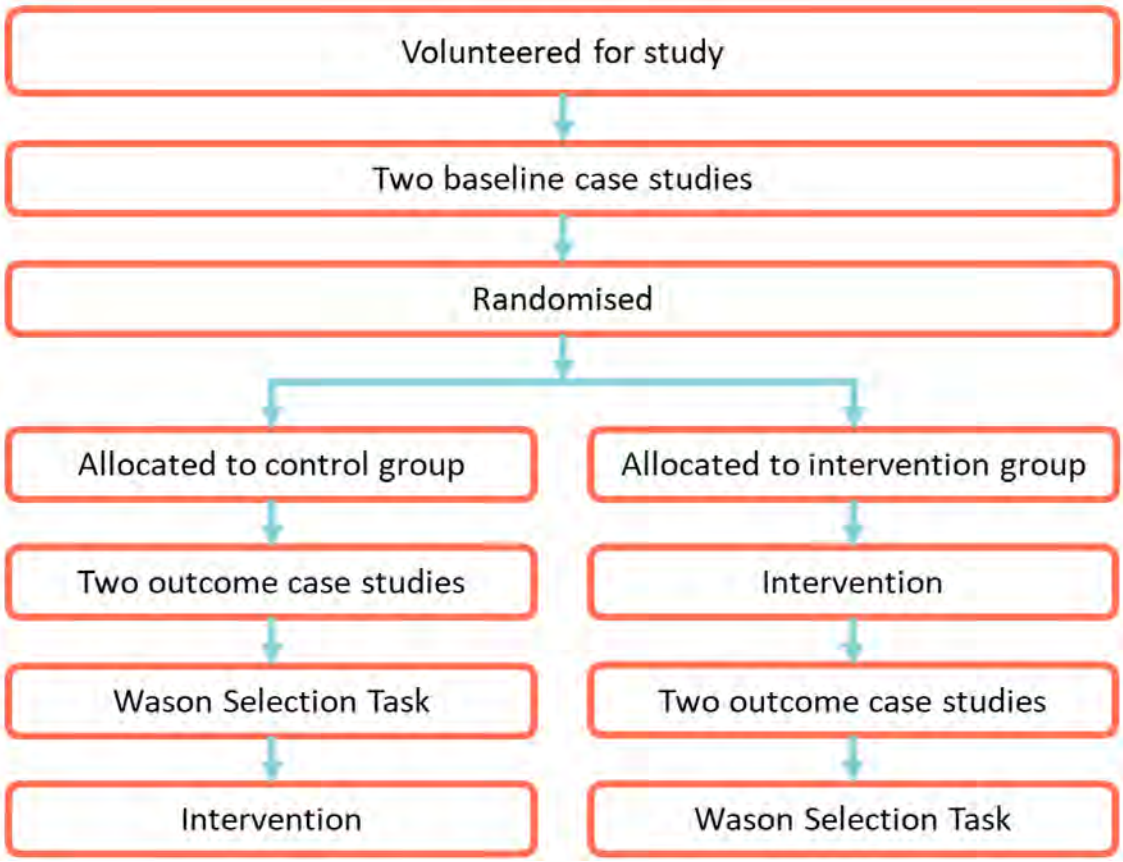


Figure 2: Diagram of crossover design

Sampling

Participants were recruited using advertisements on the What Works for Children's Social Care website and social media platforms where social workers and social work students from England were asked to take part. Partner Local Authorities and other social work organisations in England (e.g., Frontline) also helped publicise the study. The study was further publicised via Cardiff University and CASCADE's social media platforms.

All registered social workers and student social workers on social work qualifying programmes in England were eligible to take part. As part of data collection, participants were asked to self-certify that they were either a social worker or student social worker in England and any that declined to do so were exited from the survey.

Participants who completed the study were offered a £10 gift voucher as a 'thank you' for taking part. All data were collected anonymously unless participants opted to provide their email address for the purpose of sending out 'thank you' gift vouchers and to allow us to contact them for future studies. All participants were allocated an anonymous participant identifier to ensure that, where participants chose to provide their email address, it was not linked with any data they provided.

As participants signed up voluntarily and without specific targeting of particular groups, the sample is not representative of the wider social work population in England.

Data collection

The survey was hosted on Qualtrics (https://www.qualtrics.com) and available for participation between December 8th 2020 and February 15th 2021. The survey took approximately 30-60 minutes to complete. All participants were asked to complete the whole survey, with those in the intervention group completing the checklist intervention before seeing the endline measures, and those in the control group seeing the checklist intervention after seeing the endline measures.

At the start of the survey, participants were asked questions about their professional and personal characteristics. Each participant then read two baseline case studies and answered four questions for each one about the likelihood of different outcomes (resulting in 10 forecasts per case study). Table 2 below provides an example of one of the case studies from our sample (not used in the survey), including the questions posed about different outcomes, and a copy of the scale that participants were asked to use when providing their forecasts.

Following the two baseline case studies, participants were randomly allocated to the control group or the intervention group. Participants in the control group were asked to complete two endline case studies and the Wason Selection Task before accessing the checklist intervention. Participants in the intervention group were asked to complete the checklist intervention first, before completing the two endline case studies, and the Wason Selection Task. The endline case studies again involved answering four questions about each one (resulting in a further 10 forecasts per case study). Participants who completed the whole survey were asked to make 40 forecasts in total.

All of the case studies used in the study were originally selected from a list provided by one local authority in England of referrals received between January and March 2019.

Having read the relevant case files, we were able to see what happened following each referral and use this knowledge to generate the questions and evaluate the accuracy of participants' forecasts. For example, for each case study we asked how likely it was that the child would come into care within the following 12-weeks, using a scale from 0 to 100, where 0 indicates *definitely will not happen* and 100 indicates *definitely will happen*. In doing this we sought to evaluate the extent to which participants were able to make accurate predictions about what happened following the referral. Participants were not asked to make their own decision or recommendation about what should have happened following the referral, nor make any judgements about the real-life decisions

Table 2: An example of the vignettes used in this study

Referral received from the police. Prior to this referral, the family were not known to children's services.

"Salma (aged 15) has sent naked full-frontal pictures showing her face via Snapchat to an older male she knows as Amal. Salma's parents discovered the images of their daughter on her mother's phone. Due to their grave concerns, they reported the matter to the police. Salma states she has been talking to Amal for around 2 weeks and that she sent the images of herself willingly. When he requested more photos, she refused, and he blocked her. Internet safety has been discussed with parents, they acted appropriately in reporting the matter to the police. From the information available, there appears to be a risk of sexual exploitation and distribution of indecent images of a child."

1. **In response to this referral, how likely is each of the following outcomes:**

   No further action

   Social work or other form of assessment

   Emergency removal into care

   Something else

2. **Within the next 12 weeks, how likely is it that Salma will become the subject of:**

   No plan

   Child in need plan

   Child protection plan

   Looked after child plan

3. **Within the next 12 weeks, will the authority convene a strategy meeting?**

4. **Within the next 6 months, will there be a further referral about this child from any source?**

Definitely will not happen	Highly unlikely but not impossible	Very unlikely	Unlikely	Probably won't happen	May or may not happen	Could happen	Quite likely	Very likely	Highly likely but not definite	Definitely will happen
0	10	20	30	40	50	60	70	80	90	100

that had been made. The explanatory text of the survey asked respondents to forecast the likelihood of different possible outcomes, rather than to think about what they might have done with a similar referral. In other words, we were testing forecasting abilities, not the inter-rater reliability of decision-making in social work.

Outcome measures

The two endline measures we used for the study were *Brier scores* (to measure forecasting accuracy) and the *Wason Selection Task* (to measure confirmation bias).

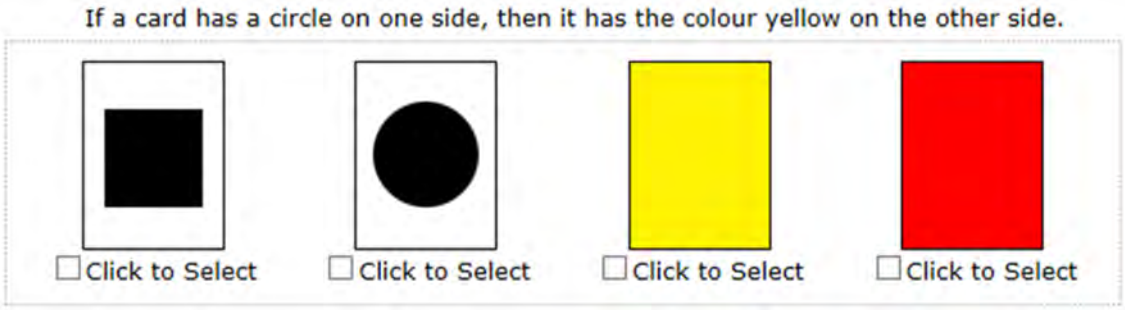
Based on responses to the case study questions, Brier scores were calculated for each participant using the following formula:

$$(1 - x)^2 + (0 - y)^2 = z$$

Where x is the probability assigned by the participant for the outcome that occurs, y is the probability assigned by the participant for the outcome that does not occur, and z is the Brier score.

As outlined in the trial protocol, Brier scores from the two baseline case studies were averaged to provide a pre-intervention Brier score. Brier scores from the two endline case studies were averaged to provide an endline Brier score. Brier scores for all four case studies were averaged to provide an overall Brier Score. Brier scores range from 0 to 2. A score of 0 represents perfect accuracy and a score of 2 represents perfect inaccuracy. A lower score indicates greater forecasting accuracy. Flipping a coin (equivalent to guessing in relation to questions with a binary outcome) would over time result in a Brier score of 0.5.

The Wason Selection Task is a logic puzzle designed to test participants' ability to identify disconfirming information in response to a series of set questions. It consists of a series of card sets and a conditional rule, such as "*if a card has a circle on one side, then it has the colour yellow on the other side*" (Figure 3). Only one side of each of the four cards is shown, and participants are asked to identify the card(s) that need to be turned over, and only that or those cards, to determine whether the rule is valid. In this study, participants were presented with three card sets. The task serves as a test of confirmation bias by measuring whether participants turn over cards that could confirm or disconfirm the rule. Participants were awarded one point for each card turned over correctly (to disconfirm the rule), and one minus point for each card turned over incorrectly (to confirm the rule). The highest possible score was +6 (indicating relatively low levels of confirmation bias) and the lowest possible score was -6 (indicating relatively high levels of confirmation bias).



The correct answer for this card set is *circle* and *red*.

Figure 3: The first item from the Wason Selection Task

Data analysis

Full details of the primary and secondary analysis are provided in the trial protocol. Participants and analysts were blind to group allocation. All data cleaning, normality assessments and analysis were carried out using the Statistical Package for the Social Sciences (SPSS) computer program Version 25 (IBM SPSS, 2017) and Microsoft Excel. Participants (n=143) were excluded from the analysis if they did not complete ≥75% of the survey. Within the remaining group of 92 participants, eleven outliers were identified using extreme values, QQ plots and boxplots. Four of the outliers were removed as their responses indicated they may not have understood the survey. The remaining outliers were included in data analysis as a comparison of original and 5% trimmed means showed the values did not strongly influence the data. All exclusions were made blind to group allocation (Figure 4).

Kolmogorov-Smirnov values indicated the assumption of normality was not violated. Histograms and Q-Q plot confirmed relatively normal distributions, however mean Brier scores were slightly positively skewed (clustered toward lower Brier score values). This is not unexpected given that the sample consisted of social workers (and one student social worker) and they were being asked to forecast outcomes within their professional domain, about which one can assume they have relevant expertise and experience.

Consort flow diagram

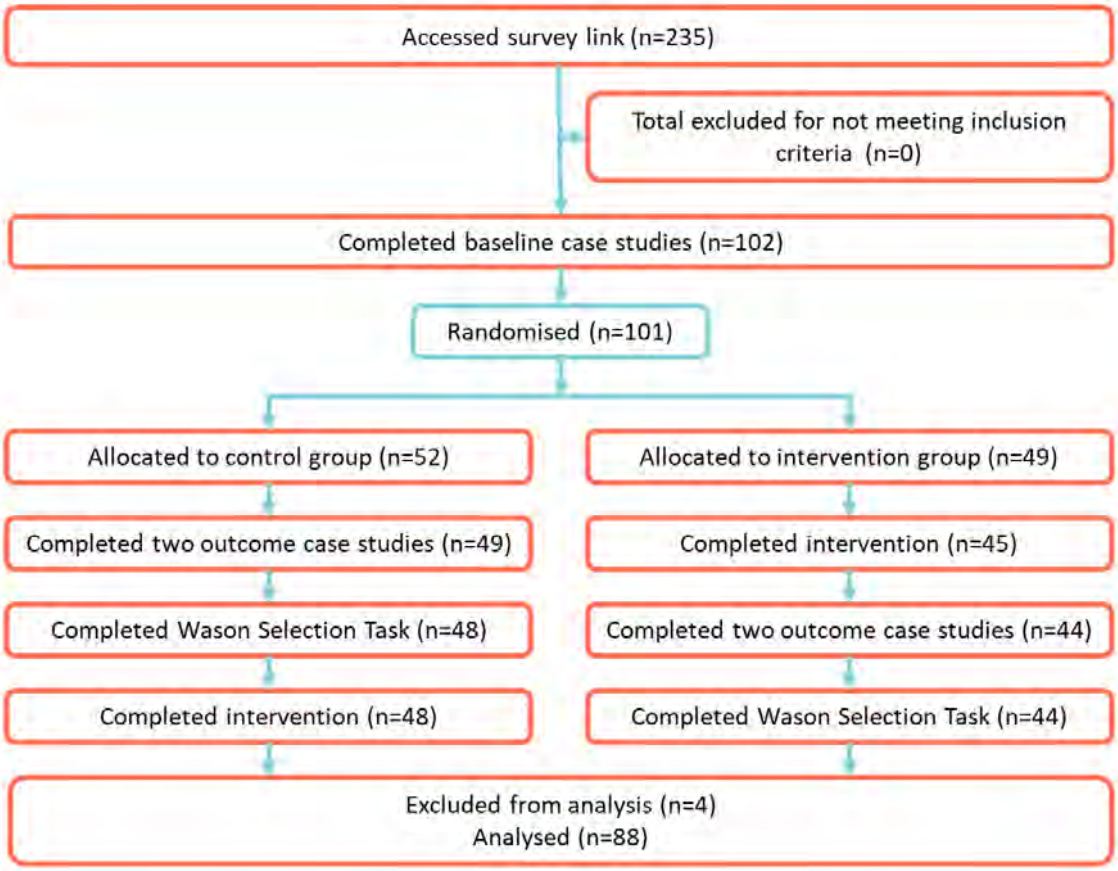


Figure 4: A consort flow-diagram overview of the sample throughout the study

FINDINGS

Sample demographics

Eighty-eight participants were included in the full analysis (meaning that we achieved 56% of our target sample size of 157 participants). The majority of participants were female, and the largest age group was 25 to 34 years old. More than a third of participants had been practicing for 10 years or more (Table 3).

Seventeen (19.3%) of the participants declared at the outset of the survey that they had taken part in previous Cardiff University studies. As some of these involved the same or similar case studies, efforts were made to change key details (e.g. names) between surveys, however there is a chance these participants had seen the case studies before and would recognise them. To determine if they performed differently from other participants, an independent samples t-test was used to compare their mean Brier scores with those of participants who had not taken part in previous studies. We found no statistically significant differences in Brier scores for participants with previous experience of similar studies (M = 0.43, SD = 0.12) compared with participants without such experience (M = 0.43, SD = 0.09), t (86) = .20, p = .84, d = 0.05 (two tailed). The mean difference in Brier scores between the two groups of participants was 0.01 with a 95% confidence interval ranging from -0.05 to 0.06. This suggests that participants who previously took part in similar studies did not perform differently compared with the other participants.

Intervention effects on forecasting accuracy

In line with the trial protocol, multiple regression was used to determine the impact of the checklist intervention on forecasting accuracy. Endline Brier scores were regressed on condition allocation, baseline Brier scores, age, gender, and length of post-qualifying experience. No assumptions were violated. The full model explained 5.2% of the variance in endline Brier scores (adjusted R2 = .05, F (5, 82) = 1.96, p = 0.09). The coefficient on the condition allocation dummy was used to determine treatment effect and indicated the checklist intervention did not have a significant impact on forecasting accuracy (*beta* = -.127, p = .23, SE = .03), with condition allocation explaining only 1.5% in the variance in endline Brier scores (Table 4 / Figure 5).

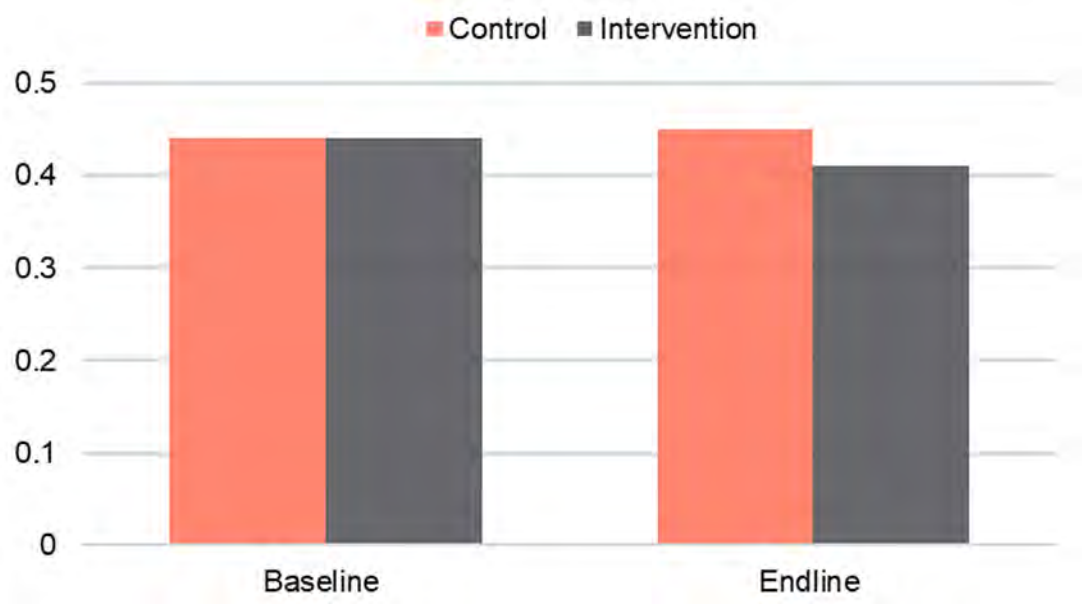
Table 3: Personal and professional characteristics of the sample (n=88)

Characteristic		N	%
Gender	Male	14	15.9
	Female	73	83.0
	Non-binary	1	1.1
Age	18 - 24	5	5.7
	25 - 34	35	39.8
	35 - 44	25	28.4
	45 - 54	11	12.5
	55 - 64	11	12.5
	65 - 74	1	1.1
Length of experience	N/A - I am not yet qualified	1	1.1
	Between 1 and 12 months	10	11.4
	Between 1 and 3 years	19	21.6
	Between 4 to 6 years	14	15.9
	Between 7 to 9 years	10	11.4
	10 or more years	34	38.6

Table 4: Baseline and endline Brier score comparison

	Control (N=44)		Intervention (N=44)	
	Mean (SD)	Median (IQR)	Mean (SD)	Median (IQR)
Baseline Brier score	.44 (.12)	.45 (.17)	.44 (.10)	.45 (.15)
Endline Brier score	.45 (.14)	.42 (.21)	.41 (.12)	.41 (.12)
Baseline-endline difference in Brier score	.01 (.02)	-0.03 (.04)	-0.03 (.02)	-0.04 (-0.03)

Participants' forecasting accuracy improved after the checklist intervention, but the change was not statistically significant



Trial of checklist intervention, n = 88

Figure 5: Comparison of Baseline and Endline Mean Brier scores

Intervention effects on confirmation bias

As per the trial protocol, multiple regression was also used to determine the impact of the checklist intervention on confirmation bias. Wason Selection Task scores were regressed on condition allocation, age, gender, and length of post-qualifying experience. No assumptions were violated. The full model explained 2.3% of the variance in Wason Selection Task Scores (adjusted R2 = .02, F (4, 82) = 1.50, p = 0.21). The coefficient on the condition allocation dummy was used to determine treatment effect and indicated the checklist intervention did not have a significant impact on confirmation bias (*beta* = -.10, p = .38, SE = .27), with condition allocation explaining only 1% in the variance in Wason Selection Task scores (Figure 6).

Relationship between forecasting accuracy and confirmation bias

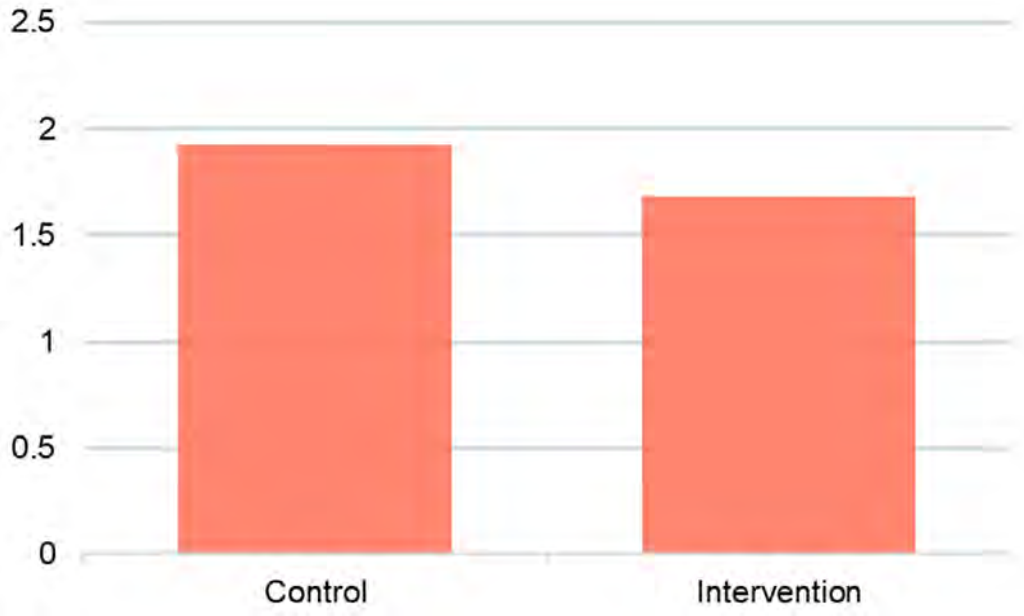
Pearson's product-moment correlation coefficient was used to explore if there was a relationship between social workers' forecasting accuracy (Brier scores) and their level of confirmation bias (Wason Selection Task scores). There was a very small negative correlation between the Brier scores and Wason Selection Task scores that did not reach statistical significance (r = -.10, n = 87, p = .37). This indicates there was no significant relationship between forecasting accuracy and confirmation bias.

Relationship between forecasting accuracy and professional / personal characteristics

The multiple regression model exploring intervention effects on forecasting accuracy was also used to determine if there was a relationship between social workers' forecasting accuracy and three personal characteristics i) age-group, ii) gender, and iii) length of post-qualifying experience.

Of the independent variables included in the model, baseline Brier scores made the strongest unique contribution to endline Brier scores and were the only significant predictor of endline Brier scores (*beta* = .225, p = .04). None of the personal or professional characteristic variables significantly predicted Brier scores (Table 5).

The checklist intervention appeared to make no meaningful difference to confirmation bias



Trial of checklist intervention, n = 88

Figure 6: Between-group comparison of Wason Selection Task scores

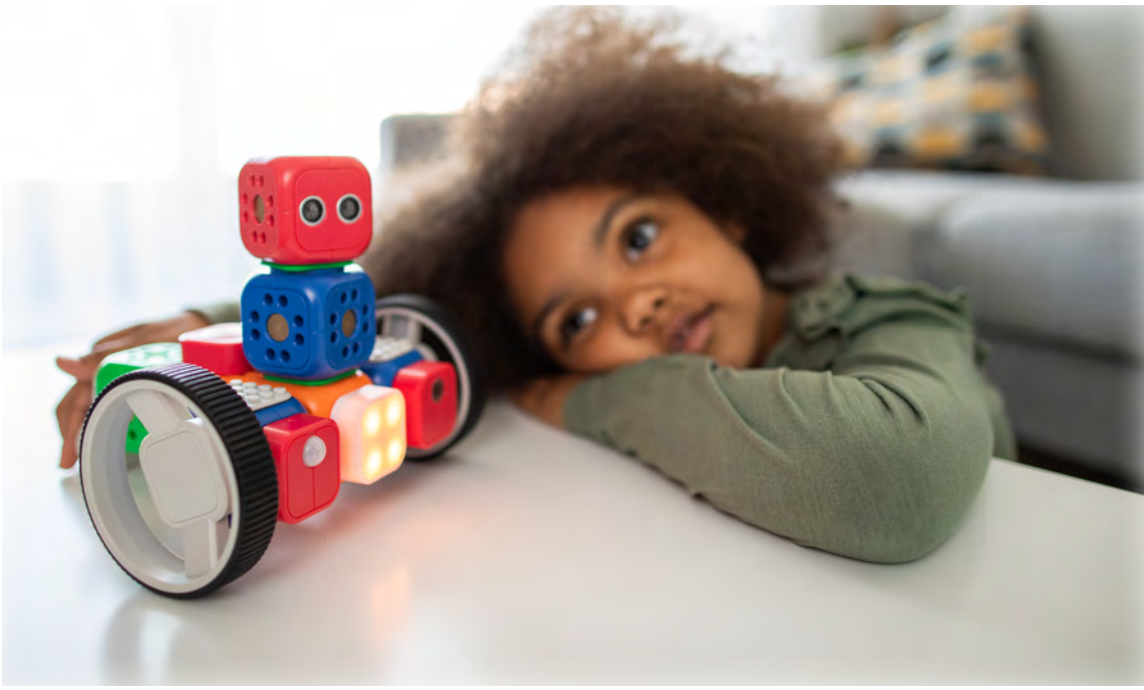
Table 5: Personal and professional characteristics and Brier scores (n=88)

None of the personal or professional characteristic variables tested significantly predicted Brier scores

Characteristic	Standardized Coefficients (Beta)	Std. Error	t value	P value	% of variance explained
Baseline Brier score	.225	.129	2.14	.035	4.9%
Gender	-.146	.037	-1.38	.172	2%
Age	.021	.014	.17	.865	0%
Length of post-qualifying experience	.141	.011	1.13	.260	1.3%

LIMITATIONS

Limitations of the study include the delivery of the intervention, the lack of a pre-intervention measure of confirmation bias, and the smaller-than-intended sample size. Checklist interventions are typically designed to be used within actual decision-making settings and are evaluated accordingly. In this study, because of the Covid-19 pandemic, we had to design an intervention that could be used online. This limits our ability to generalise from these findings, to what might happen if the same (or a similar) checklist were used in real-life decision-making environments. This is because any similar tools would most likely be used differently in practice, for example in discussion with a supervisor in a supervision session. We also used a measure of confirmation bias (the Wason Selection Test) that cannot be used in a pre- and post-test design (because we were unable to find any such measures) and as a result, we were not able to compare participants' levels of confirmation bias before and after seeing the checklist. Finally, the number of participants who completed ≥75% of the survey was lower than intended, lowering the likelihood of detecting significant effects (or non-effects) of the intervention.



DISCUSSION

Making accurate forecasts about the future, particularly within the complex environment of social work, is very challenging. Yet social workers routinely make these kinds of judgements, however implicitly, and are required by the 1989 Children Act to assess the likelihood of future significant harm in relation to children. In a previous study, we were able to identify at least some examples of these forecasts by reading assessment reports (Table 6).

Table 6: Examples of forecasts made by social workers in assessment reports (Wilkins & Forrester, 2020

1. **"I am concerned that Billy[1] is self-harming, this could impact on his health and potentially be life-threatening."**
2. **"Amber's health and development needs are not being met, which could result in her not getting the right support. Amber could have a poor start to her school life."**
3. **"Kat is not currently in education, this will impact on her achievement and future life chances, as well as her social and emotional well-being."**
4. **"The children have not been registered with a GP, and this will impact on their physical health."**
5. **"Ms Smith has been referred to attend a parenting course, which it is hoped will help her learn to manage Holly's violent behaviour."**

1 All of the examples here are pseudonyms

In relation to the current study, we found the checklist intervention did not make a significant difference for either forecasting accuracy or cognitive bias. More in-depth interventions might be required to make a meaningful difference. Another notable finding is the lack of a correlation between personal and professional characteristics and forecasting accuracy. It might be assumed that more experienced social workers would make more accurate forecasts and in some studies, less experienced workers have found it harder to analyse information (Drury-Hudson, 1999). On the other hand, Tetlock's work has shown that many experienced experts struggle to make accurate forecasts, and that some elements of greater experience,

such as being recognised as an expert, are actually negatively correlated with forecasting accuracy (Tetlock & Gardner, 2016). It may be that other characteristics, beyond those we looked at in this study, would help predict forecasting accuracy to a greater degree – for example, different ways of thinking such as being more or less open-minded or more or less comfortable with uncertainty.

Before concluding this report, it may be helpful to reflect briefly on three wider issues in relation to forecasting in social work. First, the nature of the practice environment. Making forecasts is relatively easy when conditions are stable and predictable. Predicting the weather in Hawaii does not take a great deal of meteorological expertise. Predicting the weather in a more changeable environment, such as Wales, is more difficult. The practice environment of child and family social work is, to stretch the metaphor, more like Wales than Hawaii. Yet within this complex environment, there are still things that may be easier and more difficult to forecast. For example, some of the case studies we used in this study have been consistently easier to forecast accurately compared to some of the others. This provides another explanation for the variable performance of the respondents. In reality, it may be that social workers are more likely to make contingent forecasts, than they are to make forecasts about the likelihood of change. If the current situation involving domestic abuse and substance misuse were to continue, then the child will likely suffer significant harm. However, particularly in care proceedings, the social worker will also make a forecast about the likelihood of change (because if it were anticipated that things would improve tomorrow, there would be much less need for care proceedings today).

Second, what does good forecasting performance look like in social work? As a baseline, it seems reasonable to expect that social workers, all else being equal, should be able to make forecasts more accurately than you would expect by chance. Given the complexity of the practice environment, it may be that any degree of accuracy better than chance is worth having. In a previous study (Wilkins et al., 2020) we found that a group of 283 social workers achieved an average Brier score of 0.47 (6% better than you would expect from chance). We found similar results in this study, with an average pre-intervention Brier score of 0.44 and an average endline Brier score of 0.43. Perhaps in relation to their own casework, social workers will make more accurate forecasts but currently we do not know (although we do have a study of this question currently underway in one London local authority).

Finally, what is the nature of the proposed relationship between forecasting, judgement and decision-making in social work? As indicated in Baumann et al's model (2014), the ecology of social work decision-making involves many different factors (Figure 7). The role of the individual decision-maker can only be understood alongside the influence of other factors, including external, case-related and organisational.

At the level of the individual, we find Dalgleish's model of assessment and decision-making to be a useful guide (Figure 8). In this model, various factors influence the formation of a judgement about the current situation. Depending on the nature of this judgement, a particular threshold for taking action may or may not be crossed. This helps us to understand the difference between forming a judgement and making a decision.

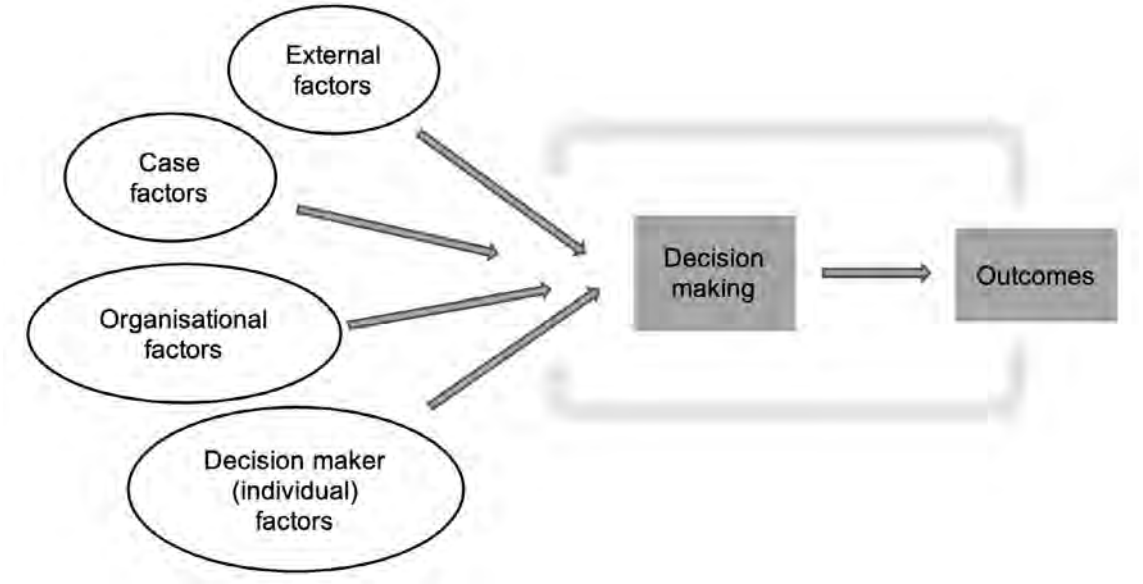


Figure 7: The ecology of social work decision-making (Baumann, Fluke, Dalgleish, & Kern, 2014)

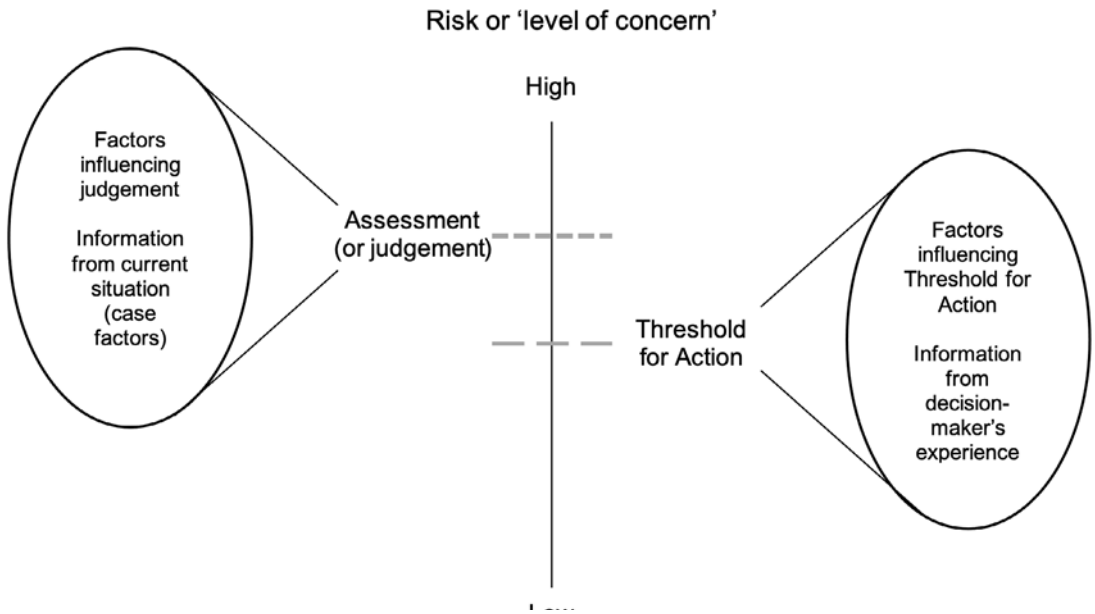


Figure 8: General Assessment and Decision-Making Model (Dalgleish, 1988)

With this model in mind, we are interested in the possible relationship between forecasting accuracy and judgement. One of our aims is to explore to what extent social workers make forecasts, and what helps promote greater accuracy. Another key aim is to test whether forecasting accuracy is a helpful proxy measure for the more nebulous concept of 'good judgement'. If so, then Brier scores could offer a useful way of testing the effectiveness of different interventions aimed at improving social work judgement (and decision-making). Finally, we are aiming in a future study to explore more explicitly the relationship between forecasting accuracy, judgement and decision-making. For example, where social workers are found to be particularly good at making forecasts, do they also show other signs of having more insight or informed judgement, and what difference does this make for the decisions they end up making?

CONCLUSIONS

Making accurate forecasts about the future, particularly within the complex environment of social work, is very challenging. Yet social workers routinely make these kinds of judgements, however implicitly, and are required by the 1989 Children Act to assess the likelihood of future significant harm in relation to children. In a previous study, we were able to identify at least some examples of these forecasts by reading assessment reports (Table 6).

Any future studies of the checklist intervention would need to ensure sufficient recruitment to enable more definitive conclusions to be drawn about its effectiveness and would also benefit from locating the study in real-life practice settings, for example in a supervision case discussion between a social worker and a manager, rather than using an online vignette design.

Future studies of interventions to improve forecasting accuracy in social work should be more in-depth than the kind of simple online interventions tested in this and a previous study (Wilkins et al., 2020). As noted in the introduction, in other fields it has been possible to test interventions and show marked improvements in forecasting accuracy as a result. These interventions tend to be much more in-depth than the checklist we tested in this study. For example, one-hour training courses can improve forecasting ability, even among those who are already proficient at forecasting (Chang, Chen, Mellers, & Tetlock, 2016). Longer training courses may be even more effective. It would also be important to move the focus of any future studies from simulated practice to real practice. This could include asking social workers to make forecasts about their own current caseload, and then waiting to judge the outcome prospectively (we have such a study currently ongoing with one London local authority).

The overall aim of this and other similar studies already completed and ongoing is to help establish what level of forecasting accuracy it is reasonable to expect from social workers, given the complexity of the practice environment, to understand more about what factors impact on their ability to make accurate forecasts, and to explore the nature of the relationship between forecasting and decision-making in social work.

REFERENCES

Baumann, Donald J, Fluke, John D, Dalgleish, Len, & Kern, Homer. (2014). The decision making ecology. *From evidence to outcomes in child welfare: An international reader*, 24-40.

Catchpole, Ken, & Russ, Stephanie. (2015). The problem with checklists. *BMJ quality & safety, 24*(9), 545-549.

Chang, Welton, Chen, Eva, Mellers, Barbara, & Tetlock, Philip. (2016). Developing expert political judgment: The impact of training and practice on judgmental accuracy in geopolitical forecasting tournaments. *Judgment & Decision Making, 11*(5).

Dalgleish, Lenard I. (1988). Decision making in child abuse cases: Applications of social judgment theory and signal detection theory. In *Advances in psychology* (Vol. 54, pp. 317-360): Elsevier.

Drury-Hudson, Julie. (1999). Decision making in child protection: The use of theoretical, empirical and procedural knowledge by novices and experts and implications for fieldwork placement. *The British Journal of Social Work, 29*(1), 147-169.

Featherston, Rebecca Jean, Shlonsky, Aron, Lewis, Courtney, Luong, My-Linh, Downie, Laura E, Vogel, Adam P, . . . Galvin, Karyn. (2019). Interventions to mitigate bias in social work decision-making: A systematic review. *Research on Social Work Practice, 29*(7), 741-752.

Gawande, Atul. (2010). *Checklist manifesto, the (HB)*: Penguin Books India.

Miller, Scott D, Hubble, Mark A, Chow, Daryl L, & Seidel, Jason A. (2013). The outcome of psychotherapy: Yesterday, today, and tomorrow.

Morewedge, Carey K, Yoon, Haewon, Scopelliti, Irene, Symborski, Carl W, Korris, James H, & Kassam, Karim S. (2015). Debiasing decisions: Improved decision making with a single training intervention. *Policy Insights from the Behavioral and Brain Sciences, 2*(1), 129-140.

Parliament. House of Lords. (2008). *Judgements – In Re B (Children) (Fc).* London: The Stationary Office.

Tetlock, Philip E. (2017). *Expert political judgment: How good is it? How can we know?-New edition*: Princeton University Press.

Tetlock, Philip E, & Gardner, Dan. (2016). *Superforecasting: The art and science of prediction*: Random House.

Westerman, Delores A. (1991). Expert and novice teacher decision making. *Journal of teacher education, 42*(4), 292-305.

Wilkins, David, Foster, Catherine, Sanders, Michael, & Reid, Louise. (2020). *Good judgement and social work decision-making: A randomised controlled trial of brief interventions to improve forecasting*. What Works for Children's Social Care.

Wilkins, David, & Forrester, Donald. (2020). Predicting the Future in Child and Family Social Work: Theoretical, Ethical and Methodological Issues for a Proposed Research Programme. *Child Care in Practice, 26*(2), 196-209.

What Works *for* Children's Social Care

CARDIFF UNIVERSITY PRIFYSGOL CAERDYDD

info@whatworks-csc.org.uk
@whatworksCSC
whatworks-csc.org.uk