The squizzel and its friends

John Paul Gosling (University of Leeds)

Squizzels

UNIVERSITY OF LEEDS

In 2004, an eminent scholar published an entertaining article in a special issue of a journal that was supposed to be about competing views on representing uncertainty.

How many legs does a squizzel have?

First tell me what a squizzel is.

Well, just use your own idea of what you think a squizzel is, and tell me how many legs it has.

……

Cooke (2004). The anatomy of the squizzel: the role of operational definitions in representing uncertainty. RESS, 85.

Squizzels

UNIVERSITY OF LEEDS

Squizzels according to Google.

Squizzels

Squizzels according to Google.

Squizzels don't exist, and their legs cannot be counted.

Squizzels

UNIVERSITY OF LEEDS

Squizzels according to Google.

Squizzels don’t exist, and their legs cannot be counted.

Squizzels exist in our minds and are not readily comparable across different individuals.

Squizzels

UNIVERSITY OF LEEDS

Probability gives us a perfect way to quantify uncertainty because it is

1) measurable,

2) comparable across different individuals,

3) able to cover all types of uncertainty.

When combining probabilities over multiple experts, weights and scores are utilised that do not exhibit properties (1) and (2).

Pooling

Let's restrict ourselves to the linear opinion pool of Stone (1961).

$$f_{pooled} = \sum_{experts} w_i f_i$$

We may believe that some experts have more to offer than others.

How can we determine the w_i in a consistent and defensible way?

Stone, M. (1961). The opinion pool. The Annals of Mathematical Statistics, 32.

Pooling

Let's restrict ourselves to the linear opinion pool of Stone (1961).

$$f_{pooled} = \sum_{experts} w_i f_i$$

We may believe that some experts have more to offer than others.

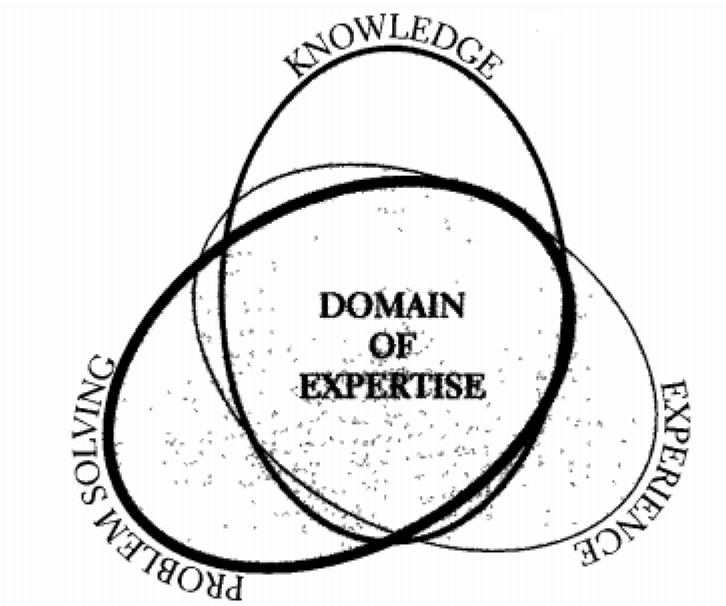
Is there an objective and unambiguous way of measuring this?

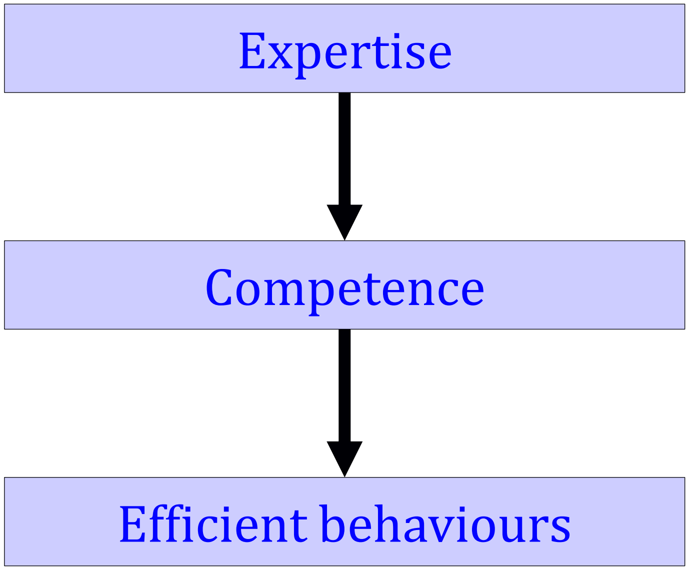
Stone, M. (1961). The opinion pool. The Annals of Mathematical Statistics, 32.

Measures of expertise

UNIVERSITY OF LEEDS

Herling (2000) gives us:





"three foundational concepts of expertise".

Herling (2000). Operational definitions of expertise and competence. Advances in Developing Human Resources, 2.

Measures of expertise

UNIVERSITY OF LEEDS

Germain and Tejeda (2012) conclude that "an instrument measuring employee expertise, as perceived by another employee, can be developed".

1. This person has knowledge that is specific to his or her field of work.
2. This person shows that they have the education necessary to be an expert in their field.
3. This person has knowledge about their field.
4. This person conducts research related to their field.
5. This person has the qualifications required to be an expert in their field.
6. This person has been trained in his or her area of expertise.
7. This person is ambitious about their work in the company.
8. This person can assess whether a work-related situation is important or not.
9. This person is capable of improving himself or herself.
10. This person is charismatic.
11. This person can deduce things from work-related situations easily.
12. This person is intuitive in their job.
13. This person is able to judge what things are important in their job.
14. This person has the drive to become what he or she is capable of becoming in their field.
15. This person is self-assured.
16. This person has self-confidence.
17. This person is an expert who is outgoing.
18. This person can talk his or her way through any work-related situation.

Generalised expertise measure

Germain & Tejeda (2012). A preliminary exploration on the measurement of expertise. Human Resource Development Quarterly, 23.

Measures of expertise

UNIVERSITY OF LEEDS

Haakma (2011) proposed using some readily-measurable quantities to derive expert weights.

Table 3 Calibration factors

Years of experience (weight 0.45)		Average number of MRI's examined per week (weight 0.45)		Examining MRI's in other areas (weight 0.1)	
X<3	1	X<5	1	X=0	1
X=>3	2	5<=X<10	2	X>0	2
		10<=X	3		

Haakma (2011). Expert elicitation to populate early health economic models of medical diagnostic devices in development. Master's thesis, University of Twente.

Measures of expertise

UNIVERSITY OF LEEDS

Haakma (2011) proposed using some readily-measurable quantities to derive expert weights.

Table 3 Calibration factors

Years of experience (weight 0.45)		Average number of MRI's examined per week (weight 0.45)		Examining MRI's in other areas (weight 0.1)	
X<3	1	X<5	1	X=0	1
X=>3	2	5<=X<10	2	X>0	2
		10<=X	3		

Naseri and Barabady (2016) were far more direct in one of their proposals:

$$w_i = \frac{y_i}{\sum_j y_j}$$

where y_i is expert i's number of years of experience.

Measures of expertise

UNIVERSITY OF LEEDS

Burgman et al. (2011) looked at some measures of expertise including an eleven point expertise rating scale.

"Participants were then asked to privately rank themselves and the other participants on an 11-point scale (0= 'no expertise', 5= 'moderate expertise', 10= 'highly expert')".

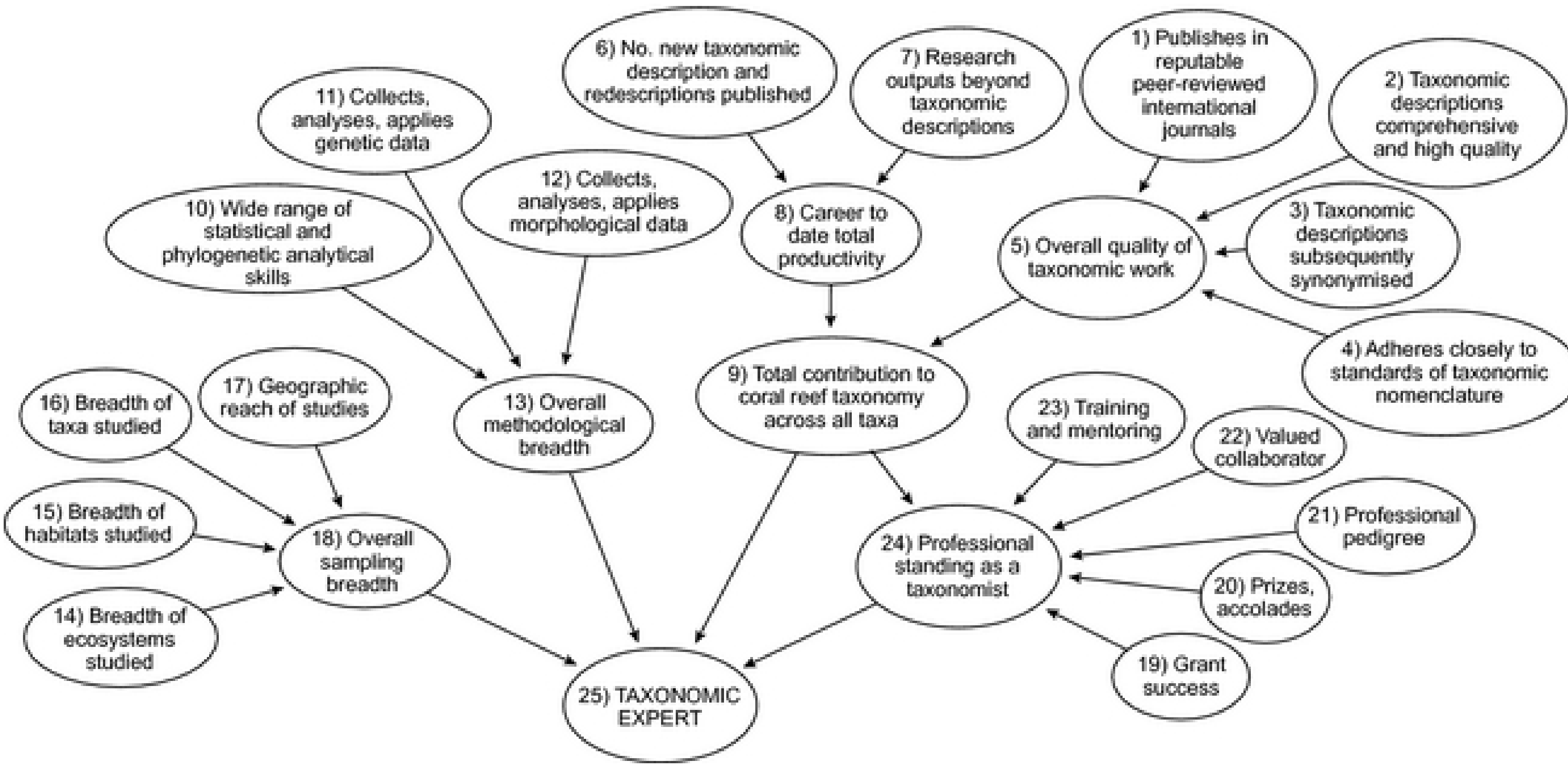
Workshop	Peer assessment versus years of experience	Peer assessment versus number of publications	Peer assessment versus qualifications	Self assessment versus peer assessment
1	**0.550 (n = 20)**	0.348 (n = 21)	**0.556 (n = 21)**	**0.675 (n = 21)**
2	**0.487 (n = 19)**	**0.587 (n = 21)**	0.064 (n = 20)	**0.684 (n = 24)**
3	**0.514 (n = 13)**	−0.123 (n = 13)	0.019 (n = 13)	**0.853 (n = 13)**
4	**0.591 (n = 25)**	**0.500 (n = 25)**	**0.489 (n = 25)**	**0.899 (n = 25)**
5	**0.836 (n = 20)**	0.309 (n = 17)	0.203 (n = 20)	**0.853 (n = 20)**
6	**0.620 (n = 14)**	0.289 (n = 14)	0.074 (n = 14)	**0.944 (n = 14)**

Burgman et al. (2011). Expert status and performance. PLoS One, 6.

Measures of expertise

UNIVERSITY OF LEEDS

Caley et al. (2014) used a Bayesian network to attempt to define expertise in a particular field.



Caley et al.(2014). What is an expert? A systems perspective on expertise. Ecology and evolution, 4.

Measures of expertise

UNIVERSITY OF LEEDS

Winkler (1968) talked about the concepts of a "better assessor" and the "goodness" of an expert, and he considered various options for weighting:

Equal,

Proportional to external ranking,

Proportional to a self-rating,

Derived from a measures of predictive performance.

But based ultimately on the judgement user's opinions.

Winkler, R. L. (1968). The consensus of subjective probability distributions. Management Science, 15.

Confidence

What is my probability of some event occurring?

$$p_1$$

Confidence

UNIVERSITY OF LEEDS

What is my probability of some event occurring?

$$p_1$$

What is my confidence in that judgement?

$$c_1$$

Confidence

UNIVERSITY OF LEEDS

What is my probability of some event occurring?

$$p_1$$

What is my confidence in that judgement?

$$c_1$$

What is my confidence in another expert's judgement (p_2)?

$$c_2$$

Confidence

UNIVERSITY OF LEEDS

What is my probability of some event occurring?

$$p_1$$

What is my confidence in that judgement?

$$c_1$$

What is my confidence in another expert's judgement (p_2)?

$$c_2$$

What is meant by "confidence"?

Capturing confidence

UNIVERSITY OF LEEDS

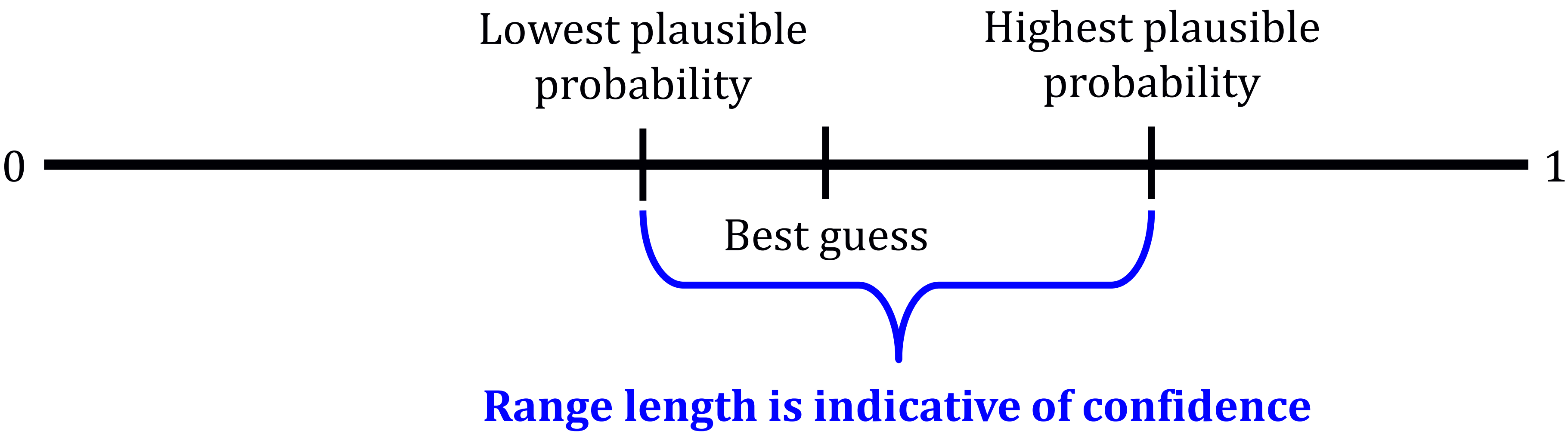
Consider a binary event.

Hanea et al. (2017). Investigate Discuss Estimate Aggregate for structured expert judgement. Int. Journal of Forecasting, 33.

Capturing confidence

UNIVERSITY OF LEEDS

Consider a binary event.

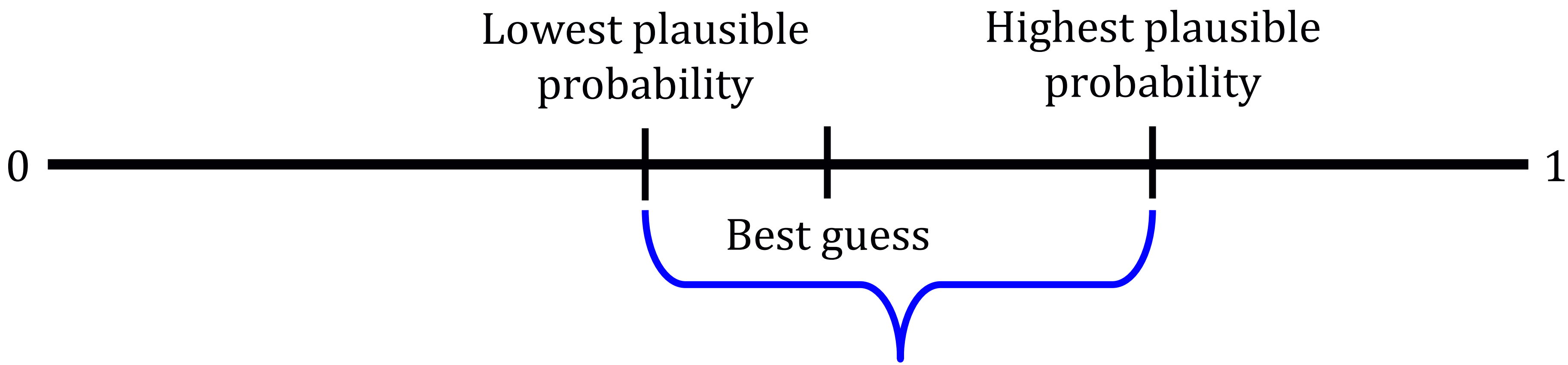


Hanea et al. (2017). Investigate Discuss Estimate Aggregate for structured expert judgement. Int. Journal of Forecasting, 33.

Capturing confidence

UNIVERSITY OF LEEDS

Consider a binary event.



Range length is indicative of confidence

What is meant by "plausible"?

And should we be using imprecise probability?

Walley, P. (1991). Statistical reasoning with imprecise probabilities.

Relative importance

UNIVERSITY OF LEEDS

DeGroot (1974) introduces a model for reaching a consensus distribution that has been revisited several times in the past 40 years.

Rounds of consensus forming are modelled by

$$\boldsymbol{F}^{(n)} = \boldsymbol{P}\boldsymbol{F}^{(n-1)} = \boldsymbol{P}^{n}\boldsymbol{F}$$

where $\boldsymbol{F}$ is a matrix encapsulating the experts' judgements about some quantity of interest and $\boldsymbol{P}$ is a matrix including weights given by the experts on the basis of relative importance of the opinions of each expert.

DeGroot (1974). Reaching a consensus. Journal of the American Statistical Association, 69.

Manipulating weights

UNIVERSITY OF LEEDS

We should worry about conflicts of interest in expert elicitation exercises.

An unscrupulous expert could adjust their weights or ratings of confidence to unduly influence the resulting distribution.

Extreme case:

$$\boldsymbol{P} = \begin{pmatrix} 0.2 & 0.4 & 0.4 \\ & & \\ & & \end{pmatrix}$$

Manipulating weights

UNIVERSITY OF LEEDS

We should worry about conflicts of interest in expert elicitation exercises.

An unscrupulous expert could adjust their weights or ratings of confidence to unduly influence the resulting distribution.

Extreme case:

$$\boldsymbol{P} = \begin{pmatrix} 0.2 & 0.4 & 0.4 \\ 0.2 & 0.7 & 0.1 \\ & & \end{pmatrix}$$

Manipulating weights

UNIVERSITY OF LEEDS

We should worry about conflicts of interest in expert elicitation exercises.

An unscrupulous expert could adjust their weights or ratings of confidence to unduly influence the resulting distribution.

Extreme case:

$$\boldsymbol{P} = \begin{pmatrix} 0.2 & 0.4 & 0.4 \\ 0.2 & 0.7 & 0.1 \\ 0 & 0 & 1 \end{pmatrix}$$

Manipulating weights

UNIVERSITY OF LEEDS

We should worry about conflicts of interest in expert elicitation exercises.

An unscrupulous expert could adjust their weights or ratings of confidence to unduly influence the resulting distribution.

Extreme case:

$$\boldsymbol{P}^n \to \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

Bayesian approaches

UNIVERSITY OF LEEDS

Moving away from linear pooling, we could model the experts' judgements as data, but the problem of relative importance does not disappear.

Bayesian approaches

Moving away from linear pooling, we could model the experts' judgements as data, but the problem of relative importance does not disappear.

Albert et al. (2012) ask each expert for a number in (0,1) that quantifies the confidence in their response.

Gelfand et al. (1995) says that we need assess the extent of knowledge about the unknown as well as sources of information and experience that have formed this knowledge.

Lindley (1983) suggests that correlation across different experts' judgements will be difficult to assess and that an inverse Wishart distribution may be inappropriate to model our uncertainty about this.

Other approaches

UNIVERSITY OF LEEDS

There have been many other attempts to rate experts based on coherence, calibration, counteracting conflicts of interest and other scoring rules…

Clemen and Winkler (1990) consider the principle of unanimity as a guide to weighting experts.

Wright et al. (1994) devised a scoring rule based upon expert coherence when assessing probabilities in different ways.

Boutilier (2012) looks at devising scoring rules that actively seek to counter experts’ conflicts through considering utilities.

Fisher et al. (2012) ask for an estimate of how sure they are that the real value lies between these bounds; here this is termed sureness and is expressed as a percentage.

Summary

UNIVERSITY OF LEEDS

If I was a decision maker, I would be worried about:

1) Ambiguities in questions posed to experts,
2) Ambiguities in interpreting experts' answers,
3) Potential for the manipulation of results.

These concerns can be avoided by using procedures that **reward honesty** and **eliminate ambiguities** (and favour people that can make probabilistic judgements).

References

UNIVERSITY OF LEEDS

Key references:

Cooke (2004). The anatomy of the squizzel: the role of operational definitions in representing uncertainty. RESS, 85.

Genest and McConway (1990). Allocating the weights in the linear opinion pool. Journal of Forecasting, 9.

Plus some that I missed off earlier slides:

Albert et al. (2012). Combining expert opinions in prior elicitation. Bayesian Analysis, 7.

Boutilier (2012). Eliciting forecasts from self-interested experts: scoring rules for decision makers. Autonomous Agents and Multiagent Systems, 2.

Fisher et al. (2012). A software tool for elicitation of expert knowledge about species richness or similar counts. Environmental Modelling & Software, 30.

Gelfand et al. (1995). Modeling Expert Opinion Arising as a Partial Probabilistic Specification. Journal of the American Statistical Association, 90.

Lindley (1983). Reconciliation of probability distributions. Operations Research, 31.

Naseri and Barabady (2016). An expert-based approach to production performance analysis… Int. Journal of System Assurance Engineering and Management, 7.

Wright, G et al. (1994). Coherence, calibration, and expertise in judgmental probability forecasting. Organizational Behavior and Human Decision Processes, 57.