## Portfolio 2. Part 1. Estimating CogSci knowledge

1. *What's Riccardo's estimated knowledge of CogSci? What is the probability he knows more than chance (0.5)?*

Riccardo’s knowledge of CogSci as estimated by using grid approximation (Figure 1a) and by using quadratic approximation (Figure 1b) is presented below. To estimate a posterior probability of a teacher giving a correct answer, a certain prior and a computed likelihood are considered. In this case, a uniform prior was elected, meaning that I expected all knowledge parameters to be equally plausible due to lack of stronger expectations. Riccardo’s estimated knowledge is visualized as a normal distribution peaking at 0.5 with a standard deviation of 0.2, which means a lot of uncertainty about plausibility of different knowledge parameters.

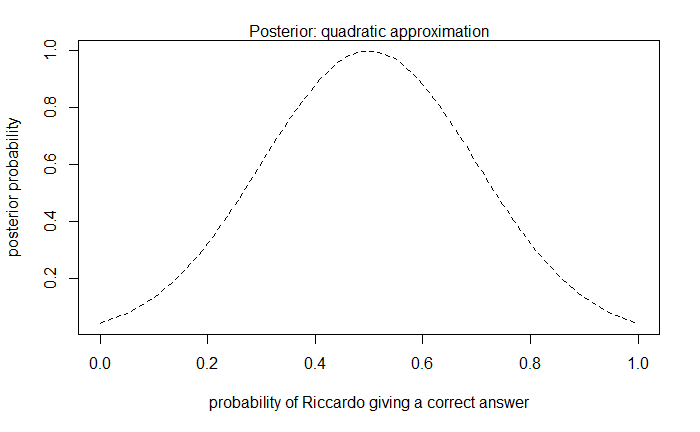
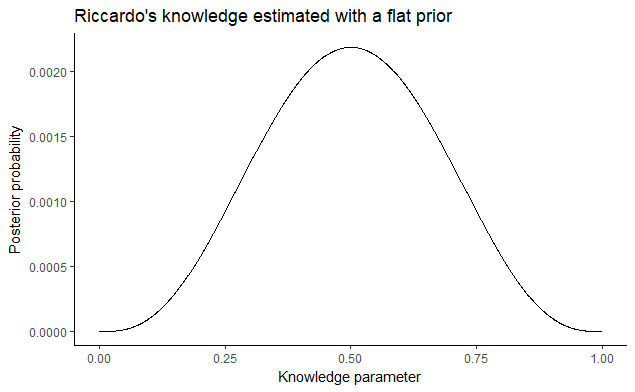


Figure 1. Posterior probability distribution estimated by a) Grid Approximation and b) Quadratic Approximation

By adding up the posterior probability where Riccardo’s knowledge is better than chance (i.e. knowledge parameter > 0.5), I found that the probability that he knows more than chance is 0.5.

1. *Estimate all the teachers' knowledge of CogSci. Who's best? Use grid approximation. Comment on the posteriors of Riccardo and Mikkel.*

*2a. Produce plots of the prior, and posterior for each teacher.*

Grid approximation method (1000 grid points) was used to estimate plausibility of different knowledge parameters (probability of a teacher giving a correct answer) for Riccardo, Kristian, Josh and Mikkel. The prior used for all teachers was a uniform prior (see Figure 2).

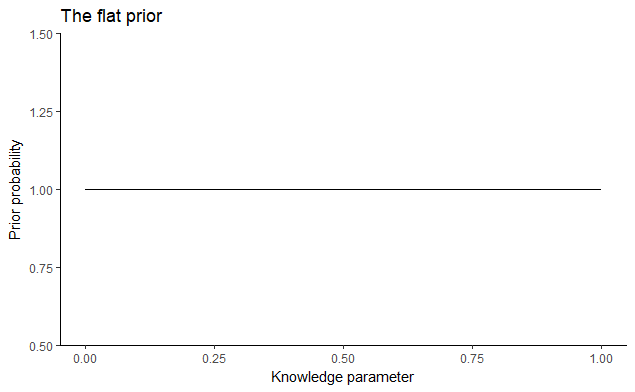
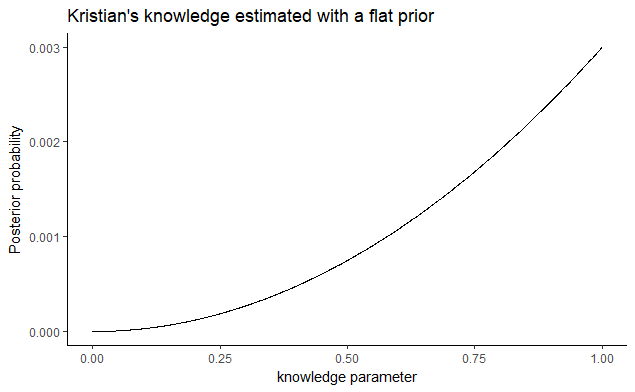
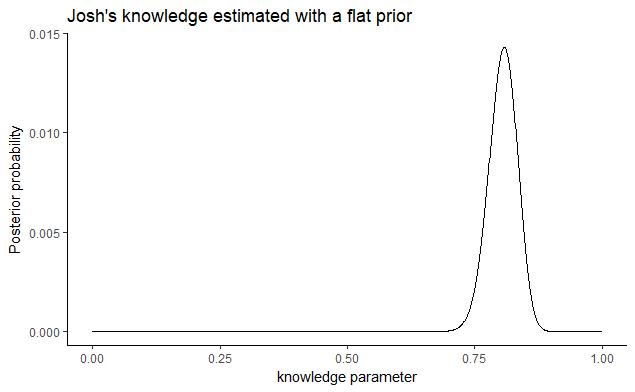


Figure 2. The uniform prior used for the grid approximation for every teacher

According to posterior distributions of knowledge parameters, that were estimated for Riccardo (Figure 1a), Kristian, Josh, and Mikkel (Figure 3 and 4), Josh showed the best knowledge of CogSci.

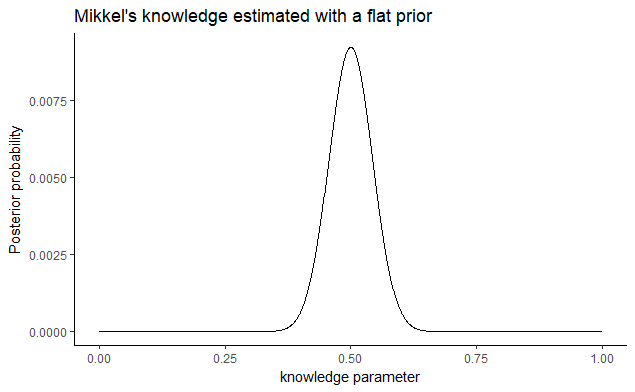


Figure 3. Posterior probability distributions of knowledge parameters for a) Kristian; b) Josh; c) Mikkel

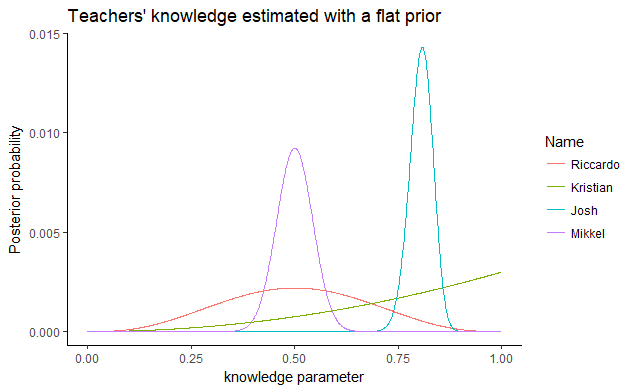


Figure 4. All teachers’ posterior distributions plotted against each other

This assessment was based on the shape of the posterior, and a calculated probability of teachers knowing better than chance.

As according to the shape of the posterior, Kristian’s knowledge parameter seems to peak at 1, but the posterior probability at that point is much lower than the probability at peaking points of other teachers’ distributions. Due to uncertainty, low knowledge parameters (below 0.5) are still plausible: the probability of him knowing better than by chance is only 0.87.

Whereas, Josh’s posterior peaks at 0.8, and even considering uncertainty (the spread around the peak), all plausible knowledge parameters are bigger than the chance probability (probability of him knowing better than by chance is 1.0).

Mikkel’s posterior distribution peaks at 0.5, just as Riccardo’s posterior distribution. However, the uncertainty about plausible knowledge parameters is lower: the spread of plausible knowledge parameters is narrower than in Riccardo’s distribution. Moreover, the posterior probability of the peaking point is much higher than the probability of the same parameter in Riccardo’s distribution. Due to a bigger amount of data taken into consideration to estimate Mikkel’s knowledge, we can be more certain about Mikkel knowing CogSci for around 0.5, than we can be sure about Riccardo knowing CogSci for around 0.5. Both teachers’ probability to know better than by chance is 0.5, but Riccardo has bigger potential to know way lower than by chance, or way higher than by chance, as compared to Mikkel.

1. *Change the prior. Given your teachers have all CogSci jobs, you should start with a higher appreciation of their knowledge: the prior is a normal distribution with a mean of 0.8 and a standard deviation of 0.2. Do the results change (and if so how)?*

*3a. Produce plots of the prior and posterior for each teacher.*

The new prior is a normal distribution with a mean of 0.8 and a standard deviation of 0.2 (see Figure 5).

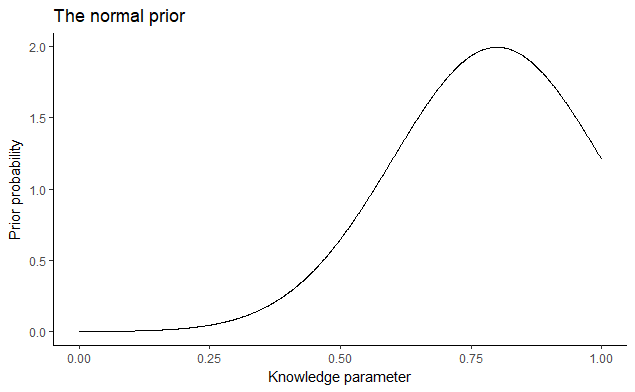


Figure 5. A normal distribution prior with a mean of 0.8 and a standard deviation of 0.2

Some of the results clearly change, as can be seen on Figure 6 (posteriors estimated with a uniform vs normal prior)

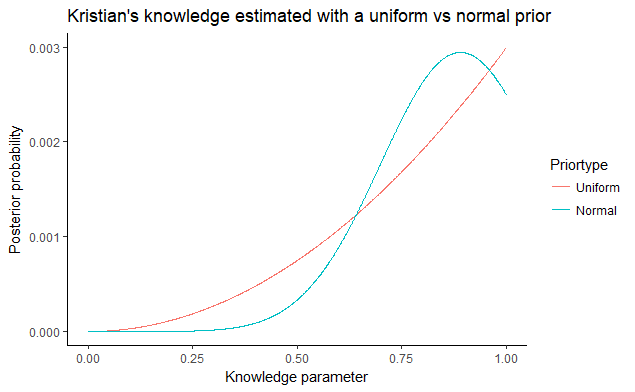
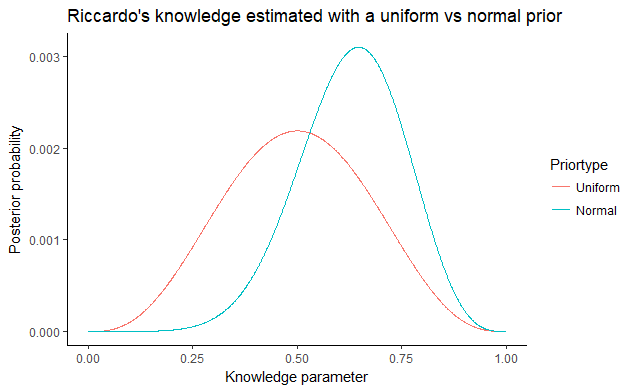
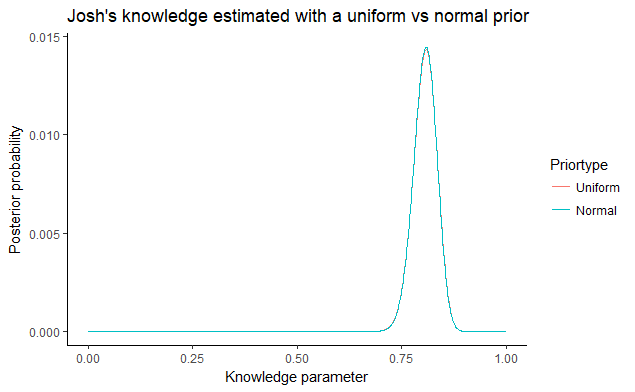
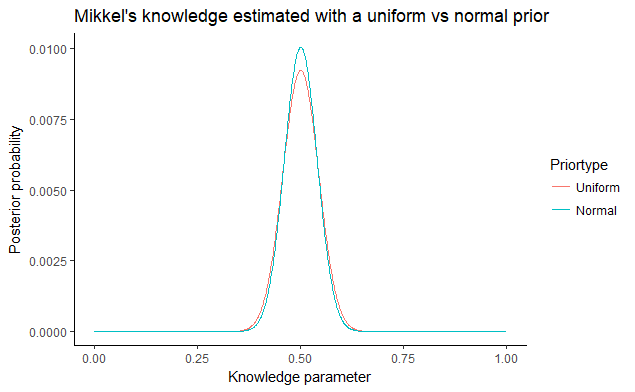
  

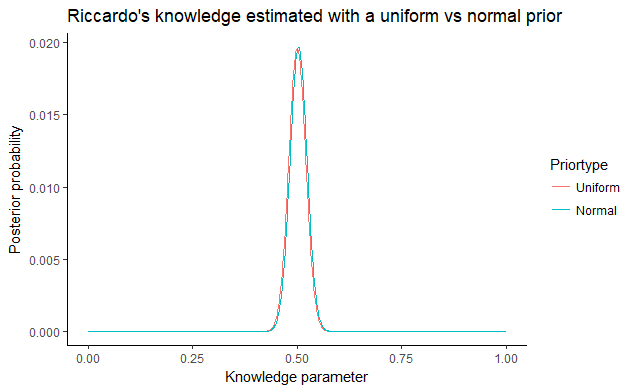
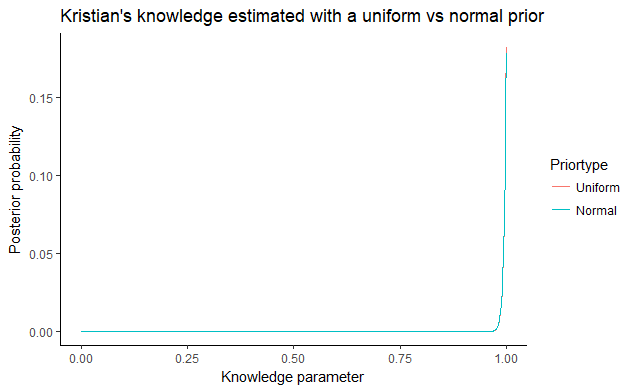
Figure 6. Posterior distributions of teachers’ knowledge parameters with a uniform prior (red color) and a normal prior (blue color)

It can be seen, that Riccardo’s and Kristian’s estimated posteriors were affected by the change of a prior way more than Josh’s and Mikkel’s posteriors. Riccardo’s posterior distribution “moved” towards the mean of the new prior and became narrower. Kristian’s new posterior probability now peaks before 1.0 and now finds very low knowledge parameters less plausible.

Riccardo’s and Kristian’s posteriors were estimated by taking into consideration less than 10 data points. The posteriors that had a lot of uncertainty about parameter probability estimation were affected by the change of a prior more, than the posteriors that had less uncertainty due to bigger amounts of data considered.

1. *You go back to your teachers and collect more data (multiply the previous numbers by 100). Calculate their knowledge with both a uniform prior and a normal prior with a mean of 0.8 and a standard deviation of 0.2. Do you still see a difference between the results? Why?*

Figure 7 shows the posteriors that were estimated with more data, and with both a uniform prior (red color) and a normal prior (blue color). Even though the proportion of correct and wrong answers remained the same, the bigger amount of data eliminated some uncertainty about the plausibility of knowledge parameters. Therefore, the priors had a smaller, if any, effect on the posterior distributions. This explains why now the posteriors were very similar regardless which priors were used to estimate them: they overlap when plotted together.

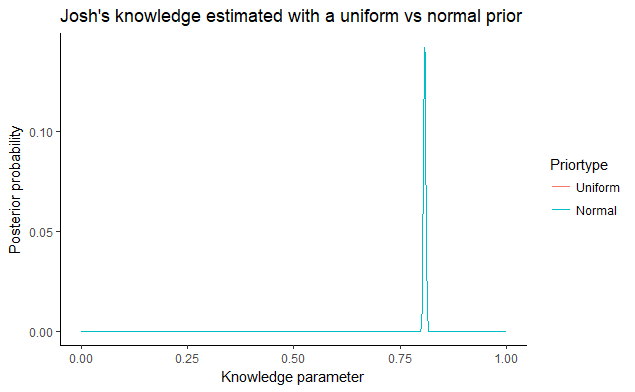
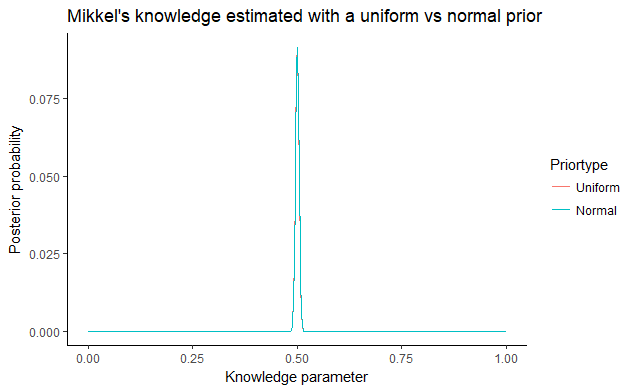
 

Figure 7. Posterior distributions of teachers’ knowledge parameters with a uniform prior in red color and a normal prior in blue color (with x100 amount of data compared to Figure 6)

1. *Imagine you're a skeptic and think your teachers do not know anything about CogSci, given the content of their classes. How would you operationalize that belief?*

To answer this question, I need to consider the design of the test (true/false answers), the scale (0 being negative knowledge, 0.5 – random chance, and 1 – awesome CogSci superpowers), and the fact that I am a skeptic. I would operationalize this belief, i.e. the prior, as a normal distribution. If the mean of this normal distribution would be equal to 0, it would mean that I expect the teachers to most likely score negative knowledge on the test, get 0 questions right, which is a very strong assumption for an unlikely event. It is more rational to assume that without knowing anything about CogSci, the teachers could still score knowledge parameters of random chance. Therefore, I would assume that the mean of the normal distribution would be 0.5. The chosen standard deviation value would define the “wiggle room” for parameters to vary. As I am supposed to be skeptic, I would try a small standard deviation value, like 0.1 (see Figure 8)

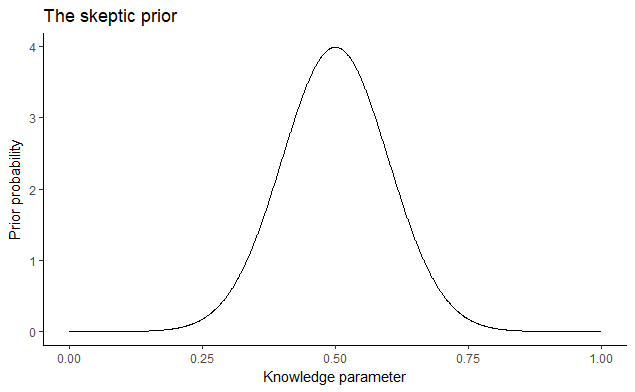


Figure 8. The skeptic prior: normal distribution with a mean of 0.5 and a standard deviation of 0.1

## Portfolio 2. Part 2. MAking Predictions

1. *Write a paragraph discussing how assessment of prediction performance is different in Bayesian vs. frequentist models*

Assessment of predictions in frequentist models (e.g. rmse on testing datasets) is very limited in its ability to inform us about its uncertainty in inference. The rmse value is calculated to quantify how close to each other sets of actual and predicted values are. The predicted values are limited to the most likely outcomes. The rmse value can’t communicate how certain the model is about plausibility of predicted values, how likely the actual data is in the world of the model, and how plausible all other possible outcomes are (like the Bayesian inference can). When assessing predictions in Bayesian models, the information about uncertainty is not lost. Therefore, it is easier to avoid overconfidence about Bayesian inference as compared to frequentist inference.

1. *Provide at least one plot and one written line discussing prediction errors for each of the teachers*.

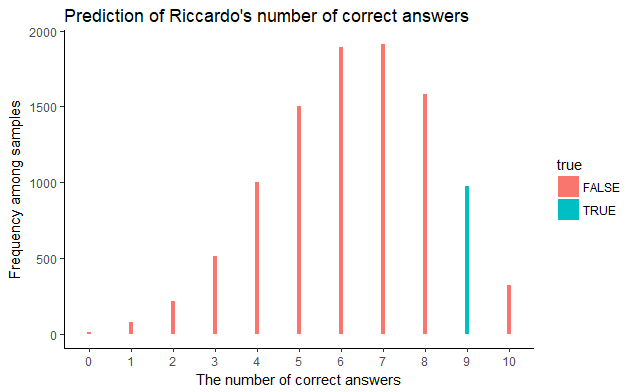
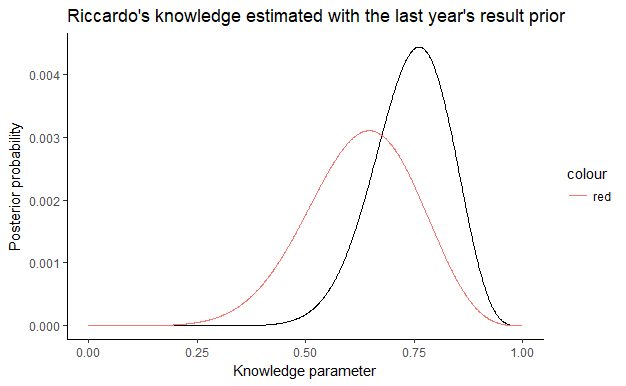


Figure 9. Way 1: the parameter estimate changed in the light of new data, it does not follow the prior. Way 2: Riccardo’s knowledge estimation produced prediction error: the actual number of correct answers (9) did not match the most likely number of correct answers according to the model (7). The model expected Riccardo to perform poorer than he did. This could either mean, that the model does not appreciate Riccardo’s knowledge enough, or the actual data is not very likely in the world of model, Riccardo could get lucky, and the model could still be right in future inference.

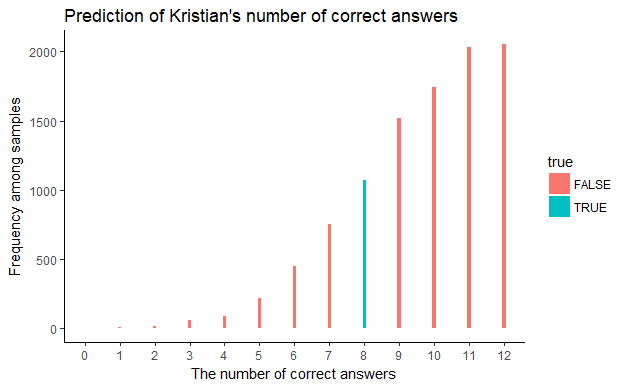


Figure 10. Kristian’s knowledge parameter estimate from the last year led to high expectations from Kristian this year. Which he did not meet. The model prediction was not accurate, the actual data and the prediction do not match.

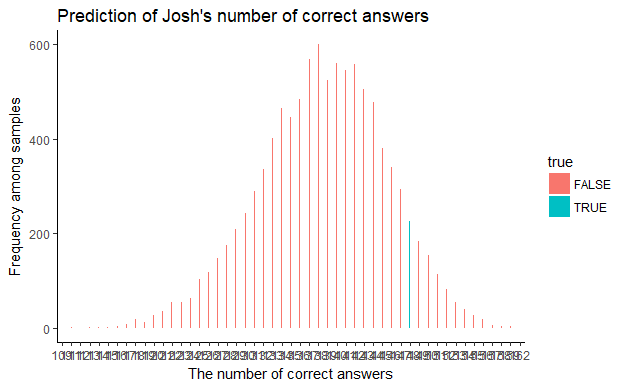


Figure 11. Josh’s performance improved, which the model did not predict. There are lots of different outcomes that the model treated as more likely than the actual data.

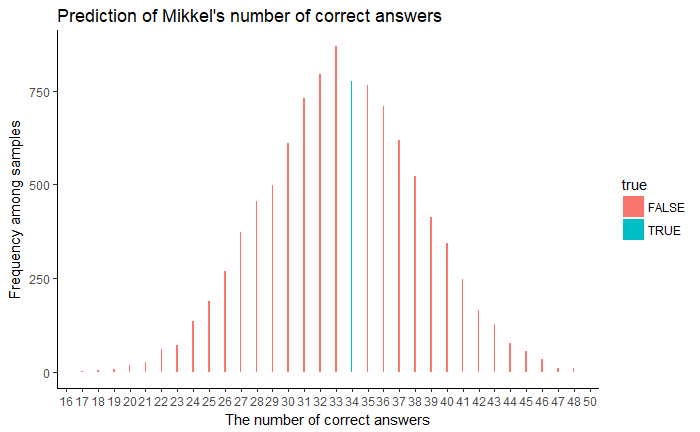


Figure 12. Mikkel’s knowledge estimate from the last year produced quite accurate predictions for this year, since the actual data matched one of the most plausible outcomes according to the model. The prediction error is the smallest in Mikkel’s model.