OSF Material 9: AIC Confidence Intervals

Table 1 displays AIC confidence intervals of the models in our initial model set (note that we used the second-order corrected version, AICc, but it equals AIC in our data due to the large sample size). Each row contains the 95% CI of the difference between the best model's AIC and the second-best model's AIC for one combination of content domain and outcome category. AIC confidence intervals were computed using the R package *nonnest2* (Merkle & You, 2016). For comparison, we printed those models bold which were included in the confidence set of models in our original analyses, that is, the models that had at least some support in the data and were thus used for inference.

The strategy that is described in the literature (Merkle, You, & Preacher, 2016) for interpreting AIC confidence intervals is the following: The AIC of the best model (i.e., the model with the smallest AIC) is tested against the AIC of the second-best model. If the AIC difference between the two models is significant (i.e., if the confidence interval excludes zero) then the best model is selected and used for further inference, while the second-best model is rejected. This was for example the case for the content domain reasoning ability and the outcome category "self-rated agentic outcomes". Here, the Beneficial PSV Model performed significantly better according to the AIC-CI than the second-best model, providing evidence for the corresponding Beneficial PSV Hypothesis. When, by contrast, the best model's AIC is not significantly different from the second-best model's AIC, this means that one cannot prefer either of the two models; evidence is inconclusive. This was for example the case for the outcome category global self-evaluation (and content domain reasoning ability): The Curvilinear PSV Model which described a beneficial but diminishing PSV effect was the best model, but it was not significantly better than the model that claims a linear beneficial PSV effect; based on the

data at hand, one cannot decide whether the beneficial PSV effect diminishes on higher PSV levels or not.

The analysis for reasoning ability and global self-evaluation (first row in Table 1) was the only analysis in which the conclusion drawn from the AIC-CI approach slightly differed from the conclusions derived in the manuscript; In our IT analyses, evidence clearly spoke for a diminishing shape of the beneficial PSV effect, while the AIC-CI approach indicated that the effect might be simply linear.

Please note that researchers who are experts in the IT approach strongly recommend against combining information-theoretic methods with significance tests (e.g., see Burnham, Anderson, & Huyvaert, 2011). The reason is that the two approaches are defined under essentially different conceptual frameworks (information-theoretic paradigm vs. null-hypothesis testing).

Table 1. *95% Confidence Intervals of AIC Difference of the Best and the Second-Best Model*

	Best model	95% CI for AIC diff.	Second-best model	Final conclusion according to the AIC-CI and according to the IT approach (main manuscript)
Reasoning ability				
Global self-evaluation	**Curvilinear PSV**	[-20.8, 7.86]	Beneficial PSV Only	Beneficial PSV Hypothesis (*manuscript*: "beneficial PSV effect curvilinear"; *AIC CI approach*: "inconclusive whether beneficial PSV effect curvilinear or linear")
Well-being	**Curvilinear PSV**	[-10.65, 1.99]	**Full model** (reflects optimal PSV level that varies for different ability levels)	Evidence for both: Beneficial PSV effect up to optimal PSV value Optimal PSV level varying for different real ability levels
Self-rated agentic outcomes	**Beneficial PSV Only**	[-36.98, -2.29]	Optimal Margin	Beneficial PSV Hypothesis
Self-rated communal outcomes	Full model not sign.			No association
Peer-rated agentic outcomes	**Beneficial PSV and Ability**	[-10.07, 4.69]	**Full model** (reflects diminishing beneficial PSV and ability effects)	Beneficial PSV effect and beneficial ability effect
Peer-rated communal outcomes	**Detrimental SE**	[-9.09, 5.66]	**Beneficial Ability Only**	Evidence for both: Detrimental SE Hypothesis Beneficial Ability Only Hypothesis
Vocabulary knowledge				
Global self-evaluation	**Full model** (reflects beneficial SE effect)	[-13.03, 10.74]	**Curvilinear PSV**	Evidence for both: Beneficial SE Hypothesis Beneficial PSV Hypothesis
Well-being	**Curvilinear PSV**	[-10.15, 5.45]	**Full model** (reflects diminishing beneficial SE effect)	Evidence for both: Beneficial PSV Hypothesis Beneficial SE Hypothesis
Self-rated agentic outcomes	**Beneficial SE**	[-89.83, -27.52]	Beneficial PSV Only	Beneficial SE Hypothesis
Self-rated communal outcomes	**Beneficial SE**	[-11.69, 9.53]	**Full model** (reflects diminishing beneficial SE effect)	Beneficial SE Hypothesis
Peer-rated agentic outcomes	**Curvilinear PSV**	[-7.71, 6.19]	**Beneficial PSV Only**	Beneficial PSV Hypothesis
Peer-rated communal outcomes	Full model not sign.			No association

Note. Best model = model with the lowest AIC. *95% CI for AIC diff.* = 95% confidence interval of the AIC difference between the best and the second-best model. Second-best model = model with the second-lowest AIC. Bold models were contained in the 95% confidence set of models (= models that were used for inference)

References

Burnham, K. P., Anderson, D. R., & Huyvaert, K. P. (2011). AIC model selection and multimodel inference in behavioral ecology: Some background, observations, and comparisons. *Behavioral Ecology and Sociobiology, 65*, 23–35. doi:10.1007/s00265-010-1029-6

Merkle, E. C., & You, D., (2016). Nonnest2: An R package for tests of non-nested models (version 0.4). Retrieved from https://cran.r-project.org/package=nonnest2

Merkle, E. C., You, D., & Preacher, K. J. (2016). Testing nonnested structural equation models. *Psychological Methods, 21*(2), 151–163. doi:10.1037/met0000038