

## SM 1.C: Exclusions due to major risk of bias

The following search results were excluded because they posed a major (i.e. extraordinary) risk of bias. While these exclusion reasons were not explicitly pre-registered, their inclusion would clearly have reduced the internal validity of our analyses.

These were based on the following criteria:

- Sources explicitly only reported significant correlations (as that will inflate the effect size estimate).
- Sources used diversity measures that correlated with team size at above  $r > .5$ , as it would then potentially serve as a proxy for team size which has well-documented relationships with team performance. (This could also apply when the correlation was not reported, but likely to be this high due to design choices.)
- Comparison groups differed substantially on an attribute other than diversity due to the study design / grouping criteria, which could introduce a plausible confound.
- Sources used performance measures that would result in inappropriate correlations for any of the following reasons:
  - Performance measures that are simple aggregates, and do not account for team size (these are then usually modelled with team size as a predictor, but their zero-order correlations cannot be meaningful).
  - Highly artificial lab-based tasks (we meant to include those by excluding student samples that did not focus on organisational tasks – but had to exclude one non-student study here).
  - Performance in a fictional world that depends primarily on the design of that fictional world (such as the diversity of fictional roles in a multiplayer computer game).
- NB: We did *not* exclude papers where diversity was closely correlated with the mean on a variable, if the authors hypothesized that diversity had an effect (given that mean levels were rarely reported, a more principled choice was not possible). However, we did exclude a paper where the hypotheses were about mean levels, with diversity as a closely correlated ( $> .5$ ) covariate.

**Table 1: Details on exclusion reasons by paper**

Reference	Reason
(Cheng et al., 2019)	<i>Non-generalizable fictional world</i> : Based on diversity of fictional roles (warrior types) in computer game world, so effect depends on game design - not necessarily informative regarding real-world diversity
(Clementi et al., 2019)	<i>Incomparable comparison groups &amp; link with group size</i> : Included single-author papers into the “homogenous” group – while they do not report a correlation between (binary) diversity and group size, it appears clear that these are thus strongly confounded.
(Congalton, 2014)	<i>Selective reporting</i> : Only reports correlations for one of two diversity variables, and merely notes that the other is not significant.
(Cummings et al., 2013)	<i>Link with group size</i> : Diversity metrics (number of disciplines/institutions) correlate very closely to team size ( $\geq .57$ ), which strongly predicts output.
(Eisenberg et al., 2021)	<i>Link with group size</i> : English language proficiency diversity of interest - yet correlates with English language proficiency at $-.74$
(Ellison & Mullin, 2014)	<i>Inappropriate performance measure</i> : Performance measure is total revenue, without accounting for drastically differing team sizes - so bivariate correlations not interpretable
(Honoré, 2022)	<i>Link with group size</i> : Diversity measures as number of founders with diverse experience - unclear diversity category, but more importantly clearly confounded with team size
(Hoover, 2017)	<i>Incomparable comparison groups</i> : Homogenous groups come from technology degrees, while mixed groups come from wide range of degrees - which will strongly confound innovation performance. Also, SD not reported so that this cannot be included without further details.
(Huber, 2017)	<i>Incomparable comparison groups</i> : Multidisciplinary teams made up of entirely different participants (young entrepreneurs, mostly graduates) than single-discipline teams (undergraduates) - so zero-order comparisons meaningless
(Jouber, 2022)	<i>Link with group size</i> : Number of women on TMT cannot be considered diversity indicator without information on TMT size
(Khan et al., 2015)	<i>Confounded with mean level</i> : Strong negative correlation between need for achievement and diversity in need for achievement ( $-.54$ ), and theoretical rationale for relationship with mean level appears stronger.

(MacCurtain, 2005)	<i>Reporting based on significance:</i> effects reported based on significance, and no response from authors to request for full table
(Marimuthu & Kolandaisamy, 2009)	<i>Excluded in spirit of 50% + 1SD rule for percentages.</i> Diversity operationalised as % of women/non-Malay managers. Figure 1 shows that both cross 50% frequently, and thus don't measure diversity (SDs not reported, so cannot apply usual decision rule)
(Mayo et al., 2020)	<i>Artificial performance task:</i> "Desert survival task" exercise - not related to organisational performance, so excluded like a student sample (even though the university lab experiment draw from a broader pool)
(Zheng et al., 2016)	<i>Non-generalizable fictional world:</i> Measures in-game performance based on variety of in-game roles represented - so measuring effect in designed rather than real world

### References

- Cheng, Z., Yang, Y., Tan, C., Cheng, D., Cheng, A., & Zhuang, Y. (2019). What makes a good team? A large-scale study on the effect of team composition in honor of kings. *The World Wide Web Conference*, 2666–2672.
- Clementi, L., Cavadini, R., Casasopra, F., Rabozzi, M., Notargiacomo, S., & Santambrogio, M. D. (2019). Diversity and inclusion: Buzzword or real value? *2019 IEEE Global Engineering Education Conference (EDUCON)*, 1228–1236.
- Congalton, J. (2014). *A mixed methods investigation of ethnic diversity and productivity in software development teams: A thesis presented in partial fulfilment of the requirements for the degree of Doctorate of Philosophy in Information Systems at Massey University, Wellington, New Zealand* [Phd]. Massey University.
- Cummings, J. N., Kiesler, S., Bosagh Zadeh, R., & Balakrishnan, A. D. (2013). Group heterogeneity increases the risks of large group size: A longitudinal study of productivity in research groups. *Psychological Science*, 24(6), 880–890.

- Eisenberg, J., Glikson, E., & Lisak, A. (2021). Multicultural virtual team performance: The impact of media choice and language diversity. *Small Group Research*, 52(5), 507–534.
- Ellison, S. F., & Mullin, W. P. (2014). Diversity, social goods provision, and performance in the firm. *Journal of Economics & Management Strategy*, 23(2), 465–481.
- Hoch, J. E., Pearce, C. L., & Welzel, L. (2010). Is the most effective team leadership shared? *Journal of Personnel Psychology*, 9(3), 105–116. <https://doi.org/10.1027/1866-5888/a000020>
- Honoré, F. (2022). Joining forces: How can founding members' prior experience variety and shared experience increase startup survival? *Academy of Management Journal*, 65(1), 248–272.
- Hoover, B. H. (2017). *Investigating if multidisciplinary or homogenous teams are more innovative in a higher education setting*. Brigham Young University.
- Huber, F. (2017). *Entrepreneurial teams and design thinking: A longitudinal evaluation of success factors* [Phd].
- Joubert, H. (2022). Women leaders and corporate social performance: Do critical mass, CEO managerial ability and corporate governance matter? *Management Decision*, 60(5), 1185–1217.
- Khan, M. S., Breitenecker, R. J., & Schwarz, E. J. (2015). Adding fuel to the fire: Need for achievement diversity and relationship conflict in entrepreneurial teams. *Management Decision*, 53(1), 75–99.
- MacCurtain, S. (2005). *An exploration of the determinants of innovation: The top management team, organisational climate and organisational learning* [Phd]. Aston University.

Marimuthu, M., & Kolandaisamy, I. (2009). *Can demographic diversity in top management team contribute for greater financial performance an empirical discussion.*

Mayo, A. T., Woolley, A. W., & Chow, R. M. (2020). Unpacking participation and influence: Diversity's countervailing effects on expertise use in groups. *Academy of Management Discoveries*, 6(2), 300–319.

Zheng, S., Zeng, X., & Zhang, C. (2016). The effects of role variety and ability disparity on virtual group performance. *Journal of Business Research*, 69(9), 3468–3477.