

Dear reviewer,

Thank you for your review and your overall positive feedback on our manuscript. Below we have added your comments in blue and the responses in black.

The paper is an important step towards standardizing techniques for the PST. Standardization bears potential to facilitate and improve future research as it makes results comparable. For practitioners the PST entails limitations due to the time consuming execution of the test. However, the findings and analysis on the test provided in this paper do hold great potential in making fracture mechanics and failure initiation more comprehensible in a teaching environment by combining empirical tests with modeling methods and offering mechanical explanations for the results.

Eventhough I believe the manuscript should be accepted as it is, I would like to offer some suggestions:

1. A more precise suggestion on what the findings indicate would be the most suitable standard. It is described that the vertical PST configuration is less susceptible to changes in the slope angle and therefore is suggested to practitioners. To my understanding of the manuscript the benefits also prevail when the crack is initiated from the uphill direction.

In general, we would have liked to have made more precise suggestions in this regard. However, it is not yet possible to say conclusively which PST geometry is best for which applications and situations. Before that, we need to understand more precisely how the slope angle affects test results. Currently, our understanding is not yet sufficient to make a concrete recommendation on PST geometry. We will therefore rewrite the section in the manuscript to discuss the pros and cons rather than making suggestions for specific user groups.

2. A topic that is touched on in the manuscript but not discussed in much detail is beam length. In the models it is assumed that the "beams were long enough, so that the tail end of the PST beam remains mechanically unchanged when the saw cut is increased and is therefore not relevant". In addition, it is mentioned that this did not apply to some results because the ratio between the depth of the weak layer and the beam length was only 0.5. In my opinion it would be interesting to discuss this issue in more depth.

We do not elaborate on this, as our study does not provide information on this aspect. Rather, we have explained the observed systematic offset of PSTs results from 20 January 2021. For more information on edge effects in PSTs, we refer to the comprehensive study of Bair et al. 2014, who discussed, among other edge effects, the influencing effect of the far end of the PST column. They concluded that crack propagation was more frequent in shorter tests due to increased stress concentration from the far edge. In their study, this edge effect occurred for PSTs up to 2 m long - or a critical cut length to column length ratio  $\geq 0.20$ . The latter ratio would therefore be more meaningful. In a revised version we will add information about the cut length of beam length ratio.