COMPARING STABILITY TESTS AND UNDERSTANDING THEIR LIMITATIONS

Karl W. Birkeland[1,2,3], Alec van Herwijnen[4], Frank Techel[4], Edward H. Bair[5,6], Benjamin Reuter[7,8], Ron Simenhois[9], Bruce Jamieson[10], Alex Marienthal[11], Doug Chabot[11], and Jürg Schweizer[4]

[1]*USDA Forest Service National Avalanche Center, Bozeman, Montana, USA*
[2]*Friends of the Colorado Avalanche Information Center, Evergreen, Colorado, USA*
[3]*Birkeland Snow and Avalanche Scientific, Bozeman, Montana USA*
[4]*WSL Institute for Snow and Avalanche Research SLF, Davos, Switzerland*
[5]*University of California Santa Barbara, Santa Barbara, California, USA*
[6]*Leidos, Inc., Reston, Virginia, USA*
[7]*Météo-France, Direction des Opérations pour la Prévision, Toulouse, France*
[8]*Centre d'Etudes de la Neige, Grenoble, France*
[9]*Colorado Avalanche Information Center, Leadville, Colorado, USA*
[10]*Department of Civil Engineering, University of Calgary, Calgary, Alberta, Canada*
[11]*Gallatin National Forest Avalanche Center, Bozeman, MT USA*

ABSTRACT: Our paper focuses on four popular snowpack tests, the Rutschblock Test (RB), Compression Test (CT), Extended Column Test (ECT), and Propagation Saw Test (PST), reviewing past and current stability test research in the context of recent advances in our understanding of dry slab avalanche release. We found that: 1) Not all tests provide the same information with regards to the critical stages of avalanche release, namely snowpack layering, failure initiation, and crack propagation, 2) Stability test research is challenging, leading to some discrepancies between studies, 3) Each stability test has its own advantages, as well as its own limitations 4) Test accuracy should not be the only selection criteria, but instead users need to pick the appropriate test to answer the questions they have about their snowpack especially with regards to failure initiation and crack propagation. Finally, conducting stability tests is valuable beyond the actual test results. Tests provide an opportunity for slowing down our thinking, focusing on the snowpack, and increasing group communication, all of which are important for minimizing common decision-making biases.

KEYWORDS: dry slab avalanche, snow stability test, snow stability, avalanche forecasting

1. INTRODUCTION

Well-placed explosives are the best method for assessing the avalanche potential of a slope (e.g., inside ski areas, above highways, etc.), but using them is not possible or practical in many situations. An alternative for gathering additional snow stability information in these cases is to conduct a stability test by isolating a block of snow to investigate some processes involved in avalanche release. However, these small-block stability tests are many orders of magnitude smaller than actual avalanches. The scale mismatch between point-scale stability tests and slope-scale avalanches means that small block tests cannot adequately capture the processes involved with avalanche release.

Recent advances in our understanding of dry-snow slab avalanche release (Schweizer et al., 2016) have sharpened our focus on the purpose of snowpack tests. Assessing snow stability at a point requires knowledge of snowpack layering (weak layer and slab properties) and the propensity for failure initiation and crack propagation (Reuter and Schweizer, 2018). In essence, we are asking: 1) Is there a slab over a weak layer?, 2) Can we initiate a failure in the weak layer?, and 3) Will the crack propagate? Crack propagation involves both the onset of crack propagation (at what critical size will a crack in a weak layer expand?) and dynamic crack propagation (how far will cracks propagate across a slope?). This latter propagation stage cannot be answered by our available stability tests.

Over three decades of spatial variability research definitively demonstrates the wide variability in slabs, weak layers, and stability test results at the slope scale (e.g., Birkeland (1990); Birkeland et al. (2004, 1995); Campbell and Jamieson (2007); Guy and Birkeland (2013); Jamieson (1995); Kronholm (2004); Kronholm and Schweizer (2003); Landry et al. (2004); Logan et al. (2004); Lutz (2009); Lutz and Birkeland (2011); Reuter et al. (2016); Schweizer et al. (2008)). All of these affect snow structure, failure initiation, and crack propagation (e.g., Kronholm and Birkeland (2005); Kronholm and Schweizer (2003)). Since pit and test locations may not be representative of trigger points on the slope, point scale assessments can only be roughly extrapolated to the slope scale. Still, results indicating instability are

* *Corresponding author address:*
Karl W. Birkeland, USDA Forest Service National Avalanche Center, Bozeman, MT 59715; tel: 406-570-7669; email: Karl.AvySci@gmail.com

clearly valuable. This is why avalanche professionals always focus on *searching for instability* when assessing avalanche potential (LaChapelle, 1980).

We define stability tests as – quite simply – snowpack tests that avalanche professionals use when assessing avalanche conditions. These tests are not used in isolation, but are one small part of the holistic approach taken by avalanche forecasters as they weigh various weather, snowpack, and avalanche data to assess avalanche likelihood. Our paper updates the review by Schweizer and Jamieson (2010) and focuses on four commonly used stability tests: 1) the Rutschblock Test (RB), 2) the Compression Test (CT), 3) the Extended Column Test (ECT), and 4) the Propagation Saw Test (PST). We briefly discuss test development before reviewing the research on test effectiveness, comparisons, and limitations. Our goal is to help avalanche professionals better understand these tests and how they can most appropriately be used for aiding their stability assessments.

2. TEST DEVELOPMENT

Here we provide a quick synopsis of the development of each test. Readers should refer to published observation guidelines (Canadian Avalanche Association, 2016; Greene et al., 2022) or one of the referenced papers for a more thorough test description. Collecting information about one or more stages of avalanche release is important (Table 1).

Table 1: Information on dry-snow slab avalanche release provided by the four snowpack tests (from Schweizer and Jamieson, 2010).

Test	Layering (slab-wkl)	Failure initiation	Crack propagation
RB (score & release type)	Yes	Yes	Yes
CT (score & FC/SQ)	Yes	Yes	Partly
ECT	Yes	Yes	Yes
PST	No	Partly	Yes

The Swiss Army first used the Rutschblock Test (RB) in the 1960s, and Föhn (1987) refined the test by identifying seven loading steps and comparing them to avalanche activity. The test involves isolating a block 2 m wide by 1.5 m upslope and progressively loading the block with a person on skis and on foot, ideally on a slope stepper than 30°. With the RB the snowpack is loaded roughly the same way as a skier-triggered slide, and it tests a larger sample than other tests. It provides good information at the relevant scale about failure initiation (score) and crack propagation (release type: whole block, most of block, or edge of block) (Greene et al., 2022; Schweizer, 2002; Schweizer et al., 1995).

The larger size of the test makes it more time consuming than the other tests, which is likely why RBs are not widely used in many countries. RBs have only been done in 1% of the 45,000+ snowpits in the SnowPilot database (note: 75% of SnowPilot pits are from the U.S. and Canada, with the remainder from various locations around the world).

Parks Canada avalanche forecasters developed the Compression Test (CT) in the 1970s (Jamieson and Johnston, 1996; Jamieson, 1999). The test involves isolating a 30 by 30 cm block of snow, placing a shovel on top of it, and then progressively tapping the shovel (10 taps from the wrist, 10 from the elbow, and the last 10 taps from the shoulder). Recent research suggests that the amount of force applied by avalanche professionals when tapping is reasonably consistent between observers (Griesser et al., 2023). The CT provides information about failure initiation, but offers little information on crack propagation propensity, since the area tested is the same as the area loaded. To address this indirectly, researchers developed two essentially equivalent systems: 1) Fracture character (Jamieson, 1999; van Herwijnen and Jamieson, 2007, 2004, 2002), and 2) Shear quality (Birkeland and Johnson, 1999; Johnson and Birkeland, 1998, 2002). The CT remains a popular test, with 57% of SnowPilot pits being associated with at least one CT.

The Extended Column Test (ECT) was developed and refined in the U.S. and New Zealand (Simenhois and Birkeland, 2009, 2006). An ECT consists of isolating a column 90 cm wide by 30 cm upslope, placing a shovel on one side of the column, tapping on the shovel with the same loading steps as the CT, and noting if a crack is initiated and whether or not it propagates completely across the column. The ECT provides an index for both failure initiation (the number of taps) and crack propagation (whether or not the crack propagates completely across the column). The ECT tests three times the surface area of the CT, but its size is still manageable, making it a relatively rapid test. However, it only tests an area about a tenth of the size of the RB. ECTs are the most widely-used test in the SnowPilot database, with at least one ECT being done in 62% of the 45,000+ total pits.

Researchers in Canada and Switzerland independently performed field experiments on crack propagation resulting in the development of the Propagation Saw Test (PST) (Gauthier and Jamieson, 2008a, 2006; Sigrist and Schweizer, 2007; van Herwijnen, 2005). The standard procedure for PSTs in North America is to isolate a column 30 cm across by 100 cm (or the weak layer depth, whichever is greater) upslope, and then cut upslope along the weak layer until the crack either propagates to the end of the column or arrests (with or without slab fracture) (Canadian Avalanche Association, 2016; Greene et al., 2022). Depending on slab thickness and critical cut length, the common length of 100 cm is too short since results are affected by edge effects (e.g., Bair et al., 2014). For the onset of crack propagation, a column length of about three to four times longer than the slab thickness or the cut length is necessary, while studies focused on slope-scale propagation have used columns as long as 10 m (Bergfeld et al., 2023b). Additionally, in contrast to the North American standards, cutting the ends of the column slope normal rather than vertical is better suited for application of the data in avalanche models; when done on slopes this configuration results in shorter critical cut lengths than vertically cut column ends (Bergfeld et al., 2023a). The PST provides a method for focusing entirely on crack propagation propensity. The tester needs to know the critical weak layer, often found by doing an ECT or CT first. The triggering mechanism is different because the critical crack length is not reached by additional loading, but instead by introducing an artificial crack. Since testers need to know the weak layer, properly prepare the column, and keep the saw exactly in the layer while cutting, the PST takes more skill than the other tests. Around 10% of SnowPilot pits are associated with a PST.

3. STABILITY TEST EFFECTIVENESS AND COMPARISONS

Evaluating and comparing the effectiveness of stability tests is challenging. Research results differ because studies use different slope stability definitions, test effectiveness metrics, and observer groups. Study locations and snowpacks also affects result. Finally, an enormous amount of well-documented spatial variability (Schweizer et al., 2008) complicates assessing test effectiveness. In the end, though test effectiveness is important, observers must choose the best option for assessing snowpack layering, failure initiation, and crack propagation for the snowpack they are assessing.

A wide variety of research has assessed the RB, CT, ECT, and PST (e.g., Birkeland and Simenhois, 2008; Föhn, 1987; Gauthier and Jamieson, 2008b, 2008a, 2006; Jamieson, 1999, 1995; Moner et al., 2008; Sigrist and Schweizer, 2007; Simenhois and Birkeland, 2009, 2006). Much work utilizes contingency tables, where assessed slope stability is compared to the test results (Table 2). Contingency tables require: 1) A stability test giving a binary result (stable/unstable), and 2) An unambiguous slope stability assessment. Both of these require judgements and assumptions. Further, some techniques for determining slope stability have relied, either implicitly or explicitly, on the stability tests being evaluated. For example, for part of their analyses, Simenhois and Birkeland (2009) relied on user-reported slope stability even though that stability was recorded after they performed their stability test. Winkler and Schweizer (2008; 2009) divided slopes into stable and unstable categories based on three criteria, one of which was the profile classification (Schweizer and Wiesinger, 2009) which includes the RB result. Any reliance on test results to assess slope stability leads to bias, a fact readily acknowledged by both of these studies.

Table 2: Researchers often use contingency tables to assess stability tests. The performance of the test (unstable/stable result) is compared against the so-called "true" stability of the snowpack (unstable/stable slope). A variety of metrics (see below table) can be calculated from these different categories of correct predictions and misses. These metrics can help us deal with unbalanced datasets and identify the source of errors, such as a test with a high false stable rate may not be a desirable test.

		Slope stability	
		Unstable	Stable
Stability test result	Unstable	True unstable	False unstable
	Stable	False stable	True stable

POD (Probability of detection) = (True unstable)/(Total unstable slopes)
PON (Probability of a null event) = (True stable)/(Total stable Slopes)
FAR (False Alarm Ratio or False Stables) = (False stable)/(Total Unstable Slopes)
POFD (Probability of False Detection) = (False unstable)/(Total Stable Slopes)
UAA (Unweighted average accuracy) = (PON + POD)/2
TSS (True Skills Score) = POD – POFD

Schweizer and Jamieson (2010) addressed some inconsistencies between studies by applying the same analyses to several original datasets. They found a higher unweighted average accuracy (UAA) for the ECT than the other tests, but they also found that comparisons within specific datasets suggested reasonably comparable results for the ECT, PST, and RB, with the CT being less accurate. Ross and Jamieson (2012) found the ECT had a higher True Skill Statistic (TSS) than the other tests.

Techel et al. (2016) compared the ECT to the RB and used an unbiased technique for rating slope stability based entirely on obvious signs of instability. They combined the RB degree and release type, using them to categorize results as unstable, intermediate, or stable, and conducted two ECTs each time. With regard to repeatability of side-by-side ECTs in the same snow pit, they reported in 21% of cases that one ECT propagated and the other did not, thus yielding contradictory results. When comparing ECTs to RBs (not considering the contradictory ECT pairs), they reported an accuracy of 80% for the RB, 72% for two ECTs, and 68% for one ECT. While the greater test accuracy of the RB compared to ECT pairs was not significant (p=0.09), it was significant compared to single ECT (p<0.01). More recently, Techel et al. (2020b) developed a four-class stability rating for the ECT by comparison with the existing RB classification scheme. They analyzed a large Swiss dataset, and found – in general – that the RB discriminates somewhat better than the ECT. They suggested that the crack propagation propensity is the most relevant ECT result and that the loading step required to initiate the failure is less indicative for stability assessment.

Marienthal et al. (2023) analyzed a dataset collected entirely by avalanche professionals (n = 561) and the larger SnowPilot dataset (n = 3,313) to compare ECTs, PSTs, and CTs. They found that ECTs and PSTs performed similarly for avalanche professionals, and that CTs had a much lower unweighted average accuracy due to excessive false unstable results. With the larger and more diverse SnowPilot data, the ECT performed better than the PST. This may be because the broader pool of SnowPilot users is less proficient at conducting PSTs, thereby negatively affecting PST performance.

Clearly, the combination of various error sources, varying definitions of slope stability, and differences in snowpacks, make it challenging to directly compare results from different studies, making within-study comparisons the most valuable for assessing test differences (Techel et al., 2020b). Defining a "best" test may not be possible, but the research suggests that the RB, ECT, and PST (and, to a lesser extent, the CT) are all valuable tests as long as observers fully understand their limitations.

4. STUDY AND TEST LIMITATIONS

Almost all stability test research points out several possible error sources, including potential misclassifications of slope stability, the relatively crude field methods, challenges identifying the most critical failure layer, and the spatial variations in test results caused by variations in slabs and weak layers (e.g., Schweizer and Jamieson (2010); Techel et al. (2020b)). Other limitations affect the various tests differently.

Minimum slab depths – For valid RB results the skier penetration must be less than the slab depth. None of the other tests have a specific minimum depth requirement. Some work on soft slabs suggested ECTs would not be effective with slab depths < 30 cm (Ross and Jamieson, 2008), but other research documents that ECTs are effective for quite shallow slabs, especially if the slabs are hard (Simenhois and Birkeland, 2009; Winkler and Schweizer, 2008). Hoyer et al. (2016) analyzed over 5000 ECTs and found that 25% of ECTPs had slab depths < 30 cm. Ultimately, the minimum slab depth is a function of the properties of the slab and weak layer.

Maximum slab depths – It may be difficult to initiate a failure in the CT, ECT, and RB when slab depths exceed about 1 m, though the exact maximum depth will depend on slab characteristics. Some work suggested that the upper limit of slab depths for ECTs is 70 cm (Ross and Jamieson, 2008), but other researchers reported valid results up to 1 m or even 1.2 m depending on the characteristics of the slab (Hoyer et al., 2016). The PST is a good option for evaluating the propagation propensity of weak layers buried so deeply that they are difficult to crack with the other tests.

Intermediate results - Simenhois and Birkeland (2009) suggested that the key piece of information from an ECT is whether or not it propagates, a result confirmed by Techel et al. (2020b). Others suggested that the applicability of the ECT could benefit from the introduction of intermediate results (Techel et al., 2020a, 2016; Winkler and Schweizer, 2009). Techel et al. (2020b) proposed categories that provide a more nuanced view of test results (Figure 1). Often, however, simply differentiating between an ECTP and an ECTN will be indicative enough. Marienthal et al. (2023) did not find an appreciable difference in the performance of the ECT when comparing Techel et al.'s (2020b) categories to the simpler ECTP vs ECTN. The PST, currently graded as stable/unstable based on whether the critical cut length is less than half the column length, might also benefit from some intermediate categories.

Slope angle dependence – Field-based research shows that ECT and CT results do not change dramatically with changing slope angle as long as the snowpack does not differ (Bair et al., 2012; Birkeland et al., 2014, 2010; Heierli et al., 2011). Similarly, cut lengths for PSTs with vertically cut column ends do not vary much with changing slope angle (Gauthier and Jamieson, 2014), but if column ends are cut slope normal then results will vary with different slope angles due to the increasing amount of snow overhanging the crack tip (Bergfeld et al., 2023). RB research suggests a modest slope angle effect (Jamieson and Johnston, 1993), though no adjustment for slope angle is needed when tests are performed on 30-40° degree slopes (Schweizer, 2002).

Time required to conduct a test – The RB test requires more digging than the other tests. ECTs, PSTs, and CTs are all reasonably fast, especially when cut with cords. PSTs conducted on deeply buried weak layers take more time, both because of the depth and the additional column length required, but they may be the only good option for testing those deep layers. Since conducting multiple tests can help to assess the spatial variations of weak layer and slab properties, several rapid tests may be more useful than one slower one.

Prior knowledge of the critical weak layer – The RB, CT, and ECT do not require prior knowledge of the critical weak layer. Indeed, a primary purpose of these tests is identifying critical layers. On the other hand, the PST requires the tester to know which weak layer to test, and to keep the saw exactly within that layer while cutting. PSTs require more skillful observers.

Test size – Stability tests are orders of magnitude smaller than avalanches and thus all suffer from edge effects. This is particularly true for the smaller tests, something that might be reduced by using longer columns for ECTs and PSTs (Bair et al., 2014). Bair et al. (2014) investigated beam lengths up to 7 m with ECTs and PSTs and Bergfeld et al. (2023) utilized 10 m PSTs for their research, but it's unclear if longer beam lengths could be helpful for practical field tests. Longer beams propagate to end less frequently than shorter tests when they are both conducted in the same pit (Bair et al., 2014). Bair et al. (2015) tested the usefulness of conducting 2 m ECTs side-by-side with standard ECTs. They found that both tests propagating in the same pit clearly indicated instability, but that assigning results with a standard ECTP and a 2 m ECTN into the "stable" category was subject to unacceptably high false-stable errors.

Tests ≠ Avalanches – Finally, it's important to remember the obvious: stability tests are not avalanches. Because they are small, they cannot represent all the processes occurring during avalanche release, in particular the slope-scale dynamic propagation process. From a research perspective, they can help us better understand some of the processes occurring during avalanche release (e.g., the use of PSTs for researching crack propagation propensity (Bergfeld et al., 2023b; Bergfeld et al., 2022; Bobillier et al., 2020, 2018; Gaume et al., 2018, 2015; Trottet et al., 2022; van Herwijnen et al., 2016)). From a practical perspective, stability tests provide valuable information, but results must *always* be viewed with appropriate caution and skepticism and a full understanding of test limitations.

5. CONCLUSIONS

Stability tests provide important data for stability evaluations during times of conditional stability. However, no tests provide a definitive "go/no go" result. With accuracies of around 80%, tests are obviously not reliable enough to bet your life on them. For all tests, some stable results are still associated with avalanche activity. Thus, assessing avalanche likelihood requires a holistic approach that integrates test results with snowpack, weather, and avalanche data. If obvious signs of instability exist – such as recent avalanches, propagating cracks, or whumpfing – stability tests are unnecessary (except to check layering) since the snowpack is clearly unstable.

We focused on four popular snowpack tests, the Rutschblock Test (RB), Compression Test (CT), Extended Column Test (ECT), and Propagation Saw Test (PST). Each test has their own strengths. While the RB and the ECT are the only two tests that provide information on layering, failure initiation and crack propagation, the PST is primarily indicative of crack propagation. The CT is well suited to find weak layers, but crack propagation propensity can only be indirectly assessed by fracture character or shear quality. Test accuracies vary substantially between studies, but ECT, PST and RB are in a similar performance range, while the CT is somewhat less accurate due to its high rate of false unstables. However, test accuracy should not be the only selection criteria, but instead users need to pick the appropriate test to answer the questions they have about their snowpack especially with regards to failure initiation and crack propagation. We suggest the following:

- ECTs are a good option in many situations, and are useful for both experienced and inexperienced testers. They can be done relatively quickly, interpreted fairly easily, and provide valuable information on layering, initiation, and propagation.
- If you have enough time and want a test that will give an intuitive "feel" for initiating an avalanche, a RB is a good option. RBs that include release type provide information on snowpack layering, crack initiation, and crack propagation, and they test a much larger area than other tests.
- PSTs are useful if you know the critical weak layer, you are experienced doing the tests, and you want to focus on crack propagation. They are especially useful for deeply buried weak layers that cannot be easily assessed with other tests, but in these cases you must be sure to test longer columns. Some research suggests the effectiveness of PSTs is less than other tests, but other work indicates that its effectiveness is roughly on par with the RB and ECT.
- CTs are useful for identifying potential weak layers, investigating near-surface weaknesses and new snow instabilities, and when you want a rapid test focusing on crack initiation.
- Avalanche professionals should consider using more than one type of test since each one provides somewhat different information about snowpack layering, failure initiation, and crack propagation. If several different tests indicate instability, it is more likely that the snowpack is unstable.

Finally, we must not overlook another way that stability tests are helpful. Slowing down the decision-making process helps us make better, less biased decisions (Kahneman, 2011). One excellent way to slow things down in the backcountry is to do a stability test. Thus, conducting any of these tests provides not only valuable snowpack information, but it also slows things down, improves group communication, and helps focus the group on snow stability.

ACKNOWLEDGEMENTS

Thanks to all the researchers, avalanche professionals, and recreational users who contributed to the datasets used for stability test research. Jim Earl maintains the SnowPilot database and analyzed some these data for our paper.

REFERENCES

Bair, E.H., Simenhois, R., Birkeland, K.W., Dozier, J., 2012. A field study on failure of storm snow slab avalanches. Cold Reg Sci Technol 79–80, 20–28. https://doi.org/10.1016/j.coldregions.2012.02.007

Bair, E.H., Simenhois, R., van Herwijnen, A., Birkeland, K.W., 2015. Using 2 m Extended Column Tests to assess snow stability. Cold Reg Sci Technol 120, 191–196.

Bair, E.H., Simenhois, R., van Herwijnen, A., Birkeland, K.W., 2014. The influence of edge effects on crack propagation in snow stability tests. Cryosphere 8, 1407–1418. https://doi.org/10.5194/tc-8-1407-2014

Bergfeld, B., Birkeland, K.W., Adam, V., Rosendahl, P.L., van Herwijnen, A., 2023a. The effect of propagation saw test geometries on critical cut length, in: Proceedings of the International Snow Science Workshop. Bend, OR.

Bergfeld, B., Van Herwijnen, A., Bobillier, G., Larose, E., Moreau, L., Trottet, B., Gaume, J., Cathomen, J., Dual, Jü., Schweizer, Jü., 2022. Crack propagation speeds in weak snowpack layers. Journal of Glaciology 68, 557–570. https://doi.org/10.1017/jog.2021.118

Bergfeld, B., van Herwijnen, A., Bobillier, G., Rosendahl, P.L., Weißgraeber, P., Adam, V., Dual, J., Schweizer, J., 2023b. Temporal evolution of crack propagation characteristics in a weak snowpack layer: conditions of crack arrest and sustained propagation. Natural Hazards and Earth System Sciences 23, 293–315. https://doi.org/10.5194/nhess-23-293-2023

Birkeland, K.W., 1990. The spatial variability of snow resistance on potential avalanche slopes. Department of Earth Sciences. Bozeman, Montana, USA.

Birkeland, K.W., Bair, E.N., Chabot, D., 2014. The effect of changing slope angle on compression test results, in: Proceedings of the 2014 International Snow Science Workshop. Banff, AB.

Birkeland, K.W., Hansen, H.J., Brown, R.L., 1995. The spatial variability of snow resistance on potential avalanche slopes. Journal of Glaciology 41, 183–189.

Birkeland, K.W., Johnson, R.F., 1999. The stuffblock snow stability test: comparability with the rutschblock, usefulness in different snow climates, and repeatability between observers. Cold Reg Sci Technol 30, 115–123.

Birkeland, K.W., Kronholm, K., Logan, S., 2004. A comparison of the spatial structure of the penetration resistance of snow layers in two different snow climates. Proceedings of the International Symposium on Snow Monitoring and Avalanches 3–11.

Birkeland, K.W., Simenhois, R., 2008. The extended column test: Test effectiveness, spatial variability, and comparison with the propagation saw test, in: Campbell, C., Conger, S., Haegeli, P. (Eds.), Proceedings of the 2008 International Snow Science Workshop. Whistler, B.C., pp. 401–407.

Birkeland, K.W., Simenhois, R., Heierli, J., 2010. The effect of changing slope angle on extended column test results: Can we dig pits in safer locations?, in: Osterhuber, R., Ferrari, M. (Eds.), 2010 International Snow Science Workshop. Squaw Valley, California, pp. 55–60.

Bobillier, G., Bergfeld, B., Capelli, A., Dual, J., Gaume, J., Van Herwijnen, A., Schweizer, J., 2020. Micromechanical modeling of snow failure. Cryosphere 14, 39–49. https://doi.org/10.5194/tc-14-39-2020

Bobillier, G., Gaume, J., van Herwijnen, A., Dual, J., Schweizer, J., 2018. Modeling the propagation saw test with discrete elements. Proceedings of the 2018 International Snow Science Workshop, Innsbruck, Austria.

Campbell, C., Jamieson, J.B., 2007. Spatial variability of slab stability and fracture characteristics within avalanche start zones. Cold Reg Sci Technol 47, 134–147.
Canadian Avalanche Association, 2016. Observation guidelines and recording standards for weather, snowpack, and avalanches. Revelstoke, B.C.
Föhn, P.M.B., 1987. The "Rutschblock" as a practical tool for slope stability evaluation, in: Salm, B., Gubler, H.U. (Eds.), Avalanche Formation, Movement and Effects. International Association of Hydrological Sciences, Davos, Switzerland, pp. 223–228.
Gaume, J., Gast, T., Teran, J., van Herwijnen, A., Jiang, C., 2018. Dynamic anticrack propagation in snow. Nat Commun 9. https://doi.org/10.1038/s41467-018-05181-w
Gaume, J., van Herwijnen, A., Chambon, G., Birkeland, K.W., Schweizer, J., 2015. Modeling of crack propagation in weak snowpack layers using the discrete element method. Cryosphere 1915–1932. https://doi.org/10.5149/tc-9-1915-2015
Gauthier, D., Jamieson, B., 2014. Evaluation of a prototype field test for fracture and failure propagation propensity in weak snowpack layers. Cold Reg Sci Technol 51, 87–97.
Gauthier, D., Jamieson, J.B., 2008a. Fracture propagation propensity in relation to snow slab avalanche release: Validating the propagation saw test. Geophys Res Lett 35, doi: 10.1029/2008GL034245.
Gauthier, D., Jamieson, J.B., 2008b. Predictions of the propagation saw test: Comparisons with the other instability tests at skier tested slopes, in: Campbell, C., Conger, S., Haegeli, P. (Eds.), Proceedings of the 2008 International Snow Science Workshop. Whistler, B.C., pp. 408–414.
Gauthier, D., Jamieson, J.B., 2006. Towards a field test for fracture propagation propensity in weak snowpack layers. Journal of Glaciology 52, 164–168.
Greene, E., Birkeland, K.W., Elder, K., McCammon, I., Staples, M., Sharaf, D., Trautman, S., Wagner, W., 2022. Snow, Weather, and Avalanches: Observation guidelines for avalanche programs in the United States, 4th edition.
Griesser, S., Pielmeier, C., Boutera Toft, H., Reiweger, I., 2023. Stress measurements in the weak layer during snow stability tests. Ann Glaciol 1–7. https://doi.org/10.1017/aog.2023.49
Guy, Z.M., Birkeland, K.W., 2013. Relating complex terrain to potential avalanche trigger locations. Cold Reg Sci Technol 86, 1–13.
Heierli, J., Birkeland, K.W., Simenhois, R., Gumbsch, P., 2011. Anticrack model for skier triggering of slab avalanches. Cold Reg Sci Technol 65, 372–381.
Hoyer, I., Greene, E., Chabot, D., Birkeland, K.W., 2016. Changes in extended column test results with varying depths, in: Proceedings of the 2016 International Snow Science Wrokshop. Breckenridge, CO, pp. 1302–1306.
Jamieson, B., Johnston, C., 1996. The compression test for snow stability, in: Proceedings of the 1996 International Snow Science Workshop. Banff, Alberta.
Jamieson, J.B., 1999. The compression test - after 25 years. The Avalanche Review 18, 10–12.
Jamieson, J.B., 1995. Avalanche prediction for persistent snow slabs. Department of Civil Engineering. Calgary AB, Canada.
Jamieson, J.B., Johnston, C.D., 1993. Rutschblock precision, technique variations and limitations. Journal of Glaciology 39, 666–674.
Johnson, R., Birkeland, K., 1998. Effectively using and interpreting stability tests. Proceedings International Snow Science Workshop, Sunriver, Oregon, U.S.A., 27 September-1 October 1998 562–565.
Johnson, R.F., Birkeland, K.W., 2002. Integrating shear quality into stability test results, in: Stevens, J.R. (Ed.), International Snow Science Workshop. Penticton, B.C., pp. 508–513.
Kahneman, D., 2011. Thinking, fast and slow. Penguin Books, London.
Kronholm, K., 2004. Spatial Variability of Snow Mechanical Properties with regard to Avalanche Formation. Faculty of mathematics and science, Department of geography. Zurich.
Kronholm, K., Birkeland, K.W., 2005. Integrating spatial patterns into a snow avalanche cellular automata model. Geophys Res Lett 32, doi: 10.1029/2005GL024373.
Kronholm, K., Schweizer, J., 2003. Snow stability variation on small slopes. Cold Reg Sci Technol 37, 453–465.
LaChapelle, E.R., 1980. The fundamental process in conventional avalanche forecasting. Journal of Glaciology, Vol. 26, No. 94 75–84.
Landry, C.C., Birkeland, K.W., Hansen, K., Borkowski, J.J., Brown, R.L., Aspinall, R., 2004. Variations in snow strength and stability on uniform slopes. Cold Reg Sci Technol 39, 205–218.
Logan, S., Birkeland, K.W., Kronholm, K., Hansen, K., Aspinall, R., 2004. Temporal changes of spatial variability of shear strength and stability, in: Elder, K. (Ed.), International Snow Science Workshop. Jackson, WY, USA.
Lutz, E., 2009. Spatial and temporal analysis of snowpack strength and stability and environmental determinants on an inclined, forest opening. Department of Earth Sciences. Bozeman, Montana.
Lutz, E.R., Birkeland, K.W., 2011. Spatial patterns of surface hoar properties and incoming radiation on an inclined forest opening. Journal of Glaciology 57, 355–366.
Marienthal, A., Chabot, D., Birkeland, K.W., 2023. Comparing the effectiveness of the ECT, PST, and CT for assessing snow stability, in: Proceedings of the 2023 International Snow Science Workshop. Bend, OR.
Moner, I., Gavalda, J., Bacardit, M., Garcia, C., Marti, G., 2008. Application of the field stability evaluation methods to the snow conditions of the eastern Pyrenees, in: Campbell, C., Conger, S., Haegeli, P. (Eds.), Proceedings of the 2008 International Snow Science Workship. Whistler, B.C., pp. 386–392.
Reuter, B., Richter, B., Schweizer, J., 2016. Snow instability pattern at the scale of a small basin. J Geophys Res Earth Surf 121, 257–282. https://doi.org/10.1002/2015JF003700
Reuter, B., Schweizer, J., 2018. Describing snow instability by failure initiation, crack propagation and slab tensile support. Geophys Res Lett 45, 7019–7027. https://doi.org/10.1029/2018GLO78069
Ross, C., Jamieson, J.B., 2008. Comparing fracture propagation tests and relating test results to snowpack characteristics, in: Campbell, C., Conger, S., Haegeli, P. (Eds.), Proceedings of the 2008 International Snow Science Workshop. Whistler, B.C., pp. 376–385.
Ross, C.K.H., Jamieson, B., 2012. The propagation saw test: Slope scale validation and alternative test methods. Journal of Glaciology 58, 407–416. https://doi.org/10.3189/2012JoG11J192
Schweizer, J., 2002. The Rutschblock test - Procedure and application in Switzerland. The Avalanche Review 20, 1, 14–15.
Schweizer, J., Camponovo, C., Fierz, C., Föhn, P.M.B., 1995. Skier triggered slab avalanche release - some practical implications, in: Sivardière, F. (Ed.), Les Apports de La Recherche Scientifique à La Sécurite Neige, Glace et Avalanche. Actes de Colloque, Chamonix, 30 Mai-3 Juin 1995. ANENA, Grenoble, pp. 309–315.
Schweizer, J., Jamieson, B., 2010. Snowpack tests for assessing snow-slope instability. Ann Glaciol 51, 187–194. https://doi.org/10.3189/172756410791386652

Schweizer, J., Kronholm, K., Jamieson, J.B., Birkeland, K.W., 2008. Review of spatial variability of snowpack properties and its importance for avalanche formation. Cold Reg Sci Technol 51, 253–272.

Schweizer, J., Reuter, B., van Herwijnen, A., Gaume, J., 2016. Avalanche release 101. Proceedings of the 2016 International Snow Science Workshop, Breckenridge, Colorado.

Sigrist, C., Schweizer, J., 2007. Critical energy release rates of weak snowpack layers determined in field experiments. Geophys Res Lett 34, L03502.

Simenhois, R., Birkeland, K.W., 2009. The Extended Column Test: Test effectiveness, spatial variability, and comparison with the Propagation Saw Test. Cold Reg Sci Technol 59, 210–216. https://doi.org/doi:10.1016/j.coldregions.2009.04.001

Simenhois, R., Birkeland, K.W., 2006. The extended column test: A field test for fracture initiation and propagation. Proceedings of the 2006 International Snow Science Workshop, Telluride, Colorado 79–85.

Techel, F., Birkeland, K.W., Chabot, D., Earl, J., Moner, I., Simenhois, R., 2020a. Comparing Extended Column Test results to signs of instability in the surrounding slopes - exploring a large, international dataset. The Avalanche Review.

Techel, F., Walcher, M., Winkler, K., 2016. Extended column test: Repeatability and comparison to slope stability and the rutschblock, in: Proceedings of the 2016 International Snow Science Workshop. Breckenridge, CO, pp. 1203–1208.

Techel, F., Winkler, K., Walcher, M., van Herwijnen, A., Schweizer, J., 2020b. On the stability interpretation of Extended Column Test results. Natural Hazards and Earth System Science. https://doi.org/doi.org/10.5194/nhess-2020-50

Trottet, B., Simenhois, R., Bobillier, G., Bergfeld, B., van Herwijnen, A., Jiang, C., Gaume, J., 2022. Transition from sub-Rayleigh anticrack to supershear crack propagation in snow avalanches. Nat Phys 18, 1094–1098. https://doi.org/10.1038/s41567-022-01662-4

van Herwijnen, A., 2005. Fractures in weak snowpack layers in relation to slab avalanche release. Department of Civil Engineering. Calgary.

van Herwijnen, A., Gaume, J., Bair, E.H., Reuter, B., Birkeland, K.W., Schweizer, J., 2016. Estimating the effective elastic modulus and specific fracture energy of snowpack layers from field experiments. Journal of Glaciology 62, 997–1007.

van Herwijnen, A., Jamieson, J.B., 2007. Fracture character in compression tests. Cold Reg Sci Technol 47, 60–68.

van Herwijnen, A., Jamieson, J.B., 2004. Fracture character in compression tests, in: Elder, K. (Ed.), International Snow Science Workshop. Jackson Hole, Wyoming, pp. 182–191.

van Herwijnen, A., Jamieson, J.B., 2002. Interpreting fracture character in stability tests, in: Proceedings ISSW 2002, International Snow Science Workshop. Penticton BC, Canada, 29 September-4 October 2002, pp. 514–522.

Winkler, K., Schweizer, J., 2009. Comparison of snow stability tests: Extended column test, rutschblock test and compression test. Cold Reg Sci Technol 59, 217–226.

Winkler, K., Schweizer, J., 2008. Comparison of different snow stability tests including the extended column test, in: Campbell, C., Conger, S., Haegeli, P. (Eds.), Proceedings of the 2008 International Snow Science Workshop. Whistler, B.C., pp. 393–400.