

EXTENDED COLUMN TEST: REPEATABILITY AND COMPARISON TO SLOPE STABILITY AND THE RUTSCHBLOCK

Frank Techel^{1,2*}, Matthias Walcher³ and Kurt Winkler¹

¹WSL-Institute for Snow and Avalanche Research SLF, Davos, Switzerland

²University of Zurich. Department of Geography, Zürich, Switzerland

³Klosters, Switzerland

ABSTRACT: Information on snow stability is key information when assessing the avalanche situation. However, stability tests like the Extended Column Test (ECT) or the Rutschblock test (RB) are point observations limited to small areas of the snowpack. Spatial variability of the snowpack may be considerable and thus test interpretation challenging. After 9 years of operational use, we determined the performance of the ECT. We explored snow profiles, where two ECTs and a RB were conducted. The main findings of our study are: (1) In 21% of the cases the ECT fracture propagation result could not be repeated at the scale of a snowpit. (2) The RB test detected more stable and unstable slopes correctly than a single ECT or two adjacent ECTs.

KEYWORDS: Extended Column Test, Rutschblock, test repeatability, snow stability, fracture propagation

1. INTRODUCTION

Gathering information about current snow stability is crucial when assessing the avalanche situation. However, direct signs of instability are sometimes lacking, particularly at lower danger levels. When danger signs are absent, snow stability tests may, beside snow stratigraphy, provide the only information on snow instability. In these situations, experienced recreationists and professionals may benefit the most by performing a stability test (Bellaire et al., 2010). To be of value, stability tests must provide repeatable results and detect a large proportion of unstable and stable slopes correctly (e.g. Schweizer and Jamieson, 2010).

The Extended Column Test (ECT), introduced in 2006 in North America and New Zealand (Simenhois and Birkeland, 2006), is now also widely used in Europe and has been included in the respective snow and avalanche observation guidelines (e.g. Darms et al., 2014). After nine winters using the ECT in Switzerland, we were interested to evaluate the performance in an operational setting, and to compare ECT and Rutschblock (RB; Föhn, 1987; Schweizer, 2002) test results with the estimated slope stability.

With this study we address the following questions:

- (1) How repeatable are ECT test results at the scale of a snowpit?
- (2) Does performing a second ECT aid in correctly detecting stable and unstable slopes?
- (3) How does the accuracy of ECT and RB compare?

2. DATA AND METHODS

2.1 Data

The data-set was collected mostly in an operational context during nine winters in the Swiss Alps, generally at locations above treeline and in northerly aspect slopes (NE-N-NW). The snowprofiles, ECT and RB were carried out by trained observers and SLF employees.

576 profiles with two ECTs adjacent to each other were performed in dry snow. A combination of two ECT and a RB were performed in 221 slopes, where a *stable-unstable* slope stability rating was available.

2.2 Stability test result interpretation and decisive result

The Extended Column Test and the Rutschblock test were carried out according to standard procedure (ECT: Simenhois and Birkeland (2006); RB: Darms et al., 2014).

* Corresponding author address:

Frank Techel, WSL Institute for Snow and Avalanche Research SLF, Flüelestrasse 11, CH-7260 Davos Dorf, Switzerland;
tel: +41 81 417 0349
email: techel@slf.ch

ECT fractures propagating across the entire test column (full propagation, *fp*) were considered *unfavorable*, when the propagation occurred during the same, or the next loading step following fracture initiation in the same layer (Winkler and Schweizer, 2009), regardless of failure plane depth. If more than one failure plane was noted, *fp*-fractures with the lowest score and, if the score was equal, closest to the surface were regarded to be the most relevant for stability assessment (= decisive result).

The interpretation of an ECT test pair was as follows:

- *unfavorable*, if both tests were *unfavorable*,
- *favorable*, if both tests were *favorable*
- *contradictory*, if one test was *favorable* and one *unfavorable*.

Rutschblock results were classified according to the stability interpretation operationally used in Switzerland. Both, the RB score and the RB release type were considered resulting in three stability classes *unfavorable*, *intermediate* and *favorable* (Table 1). The decisive result was the one leading to the lowest RB stability class, or if several fractures with the same class occurred, the one closest to the surface.

Tbl 1: Rutschblock stability classification. Abbreviations for the release type: wb - whole block, pb - partial break (below ski, edge only), all – any of the before-mentioned.

Stability	Score and release type	N
unfavorable	RB1-all, RB2-all, RB3+wb	65
intermediate	RB3+pb, RB4-all, RB5+wb	81
favorable	RB5+pb, RB6-all, RB7-all	75

2.3 Classification of slope stability

As in previous studies exploring the accuracy of stability tests, we classified slopes according to observations relating to snow instability, like recent avalanche activity or danger signs (whumpfs or shooting cracks). If any danger signs or recent avalanche activity (natural or skier-triggered avalanches from the day of observation or the previous day) were noted in the slope where the test was carried out or in a neighbouring slope, the slope was considered *unstable*. A slope was considered *stable*, if it was clearly stated that on the day of observation none of the before-mentioned signs were observed, neither in the tested slope

nor elsewhere. The distribution of slope stability ratings is shown in Table 2.

Tbl. 2: Slope stability rating.

N	Unstable	Stable
221	26%	74%

2.4 Statistical methods

The performance of snow stability tests was analyzed using the measures of

- sensitivity or *correct unstable* (correct detection of *unstable* slopes)
- specificity or *correct stable* (correct detection of *stable* slopes)
- unweighted average *accuracy* (the mean of sensitivity and specificity)
- *false-alarm* rate (slopes classified as *unstable* when in fact *stable*) and
- *misses* (*unstable* slopes falsely classified as *stable*).

These are described in detail in Doswell et al. (1990) or, applied to snow stability tests, in Schweizer and Jamieson (2010).

We applied the chi-square based non-parametric two proportion test without continuity correction (R Core Team, 2016) to test whether two proportions were significantly different.

Significance level: $p < 0.05$

3. RESULTS

3.1 Repeatability of ECT at the snowpit scale

576 sets of two ECTs were conducted in dry snow conditions. In 37% of these pairs (N=211) the decisive fracture was twice *unfavorable* and in 243 cases (42%) it was twice *favorable* (Table 3). *Contradictory* fracture propagation (one *favorable* and one *unfavorable*) results were noted in 21% of the cases (N=122).

Tbl. 3: Summary of fracture propagation results of ECT pairs.

N	Unfavorable	Contradictory	Favorable
576	37%	21%	42%

3.2 Comparison between slope stability and test results

Comparing the slope stability rating (*stable* and *unstable* slopes) with stability test results (*favorable* and *unfavorable* test results) showed generally greater test accuracy when using the Rutschblock (80%, not considering *intermediate* results), compared to two ECTs (72% agreement, not considering *contradictory* results) or single ECT results (68% accuracy, Table 4, Fig. 1). 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 ($p<0.01$). The RB performed significantly better in the test measures *correct unstable*, *false-alarms*, *misses* ($p<0.01$). While the *correct stable* predictions based on RB tests were also somewhat better than for single ECT or two ECT, this was not significant.

Unfavorable single ECT and *unfavorable* ECT pairs were noted more frequently in *stable* than *unstable* slopes (*false alarm* ratio for single ECT 58%, and for two ECT 53%, Table 4). *Misses - favorable* test results in *unstable* rated slopes - were also much more frequent using the ECT than the RB.

Similar to the full dataset of ECT pairs (section 3.1), 20% of the 221 ECT pairs in slopes where a slope stability rating was available, resulted in *contradictory* results. This proportion was similar in *stable* and *unstable* slopes (20% vs. 19%; Fig. 2, left).

As can be seen in Table 4, *contradictory* ECT test results can neither clearly be considered as belonging to the *stable* group, nor are they an indicator of *unstable* conditions. In both cases, the *accuracy* decreased, and the *correct unstable* – *correct stable* ratio became more unbalanced.

Intermediate RB results, which are usually interpreted as indicating *transitional* snow stability (*transitional* between *stable* and *unstable* conditions), were observed more often in *stable* slopes (43%) than *unstable* slopes (19%; Fig. 2, right). Although not shown, a considerable drop in the RB's test accuracy resulted, when *intermediate* RB stability was considered as an indicator of a *stable* slopes (77%) or *unstable* slope stability (66%).

Tbl. 4: Detection of *stable* and *unstable* slopes with a single ECT (each ECT test of an ECT test pair was compared separately with the slope stability rating), two ECT (without *contradictory* results) and RB (without *intermediate* stability classification), and when interpreting *contradictory* ECT results as *stable* or *unstable*. The results *correct unstable* and *correct stable* are presented in Figures 1 and 2.

<i>test</i>	<i>N</i>	<i>accuracy</i>	<i>correct unstable</i>	<i>correct stable</i>	<i>false-alarms</i>	<i>misses</i>
single ECT	442	0.68	0.69	0.66	0.58	0.31
two ECT _{without contradictory}	177	0.72	0.74	0.70	0.53	0.26
RB _{without intermediate}	140	0.80	0.87	0.73	0.38	0.13
two ECT _{contradictory as stable}	221	0.68	0.60	0.76	0.53	0.40
two ECT _{contradictory as unstable}	221	0.68	0.79	0.56	0.61	0.21

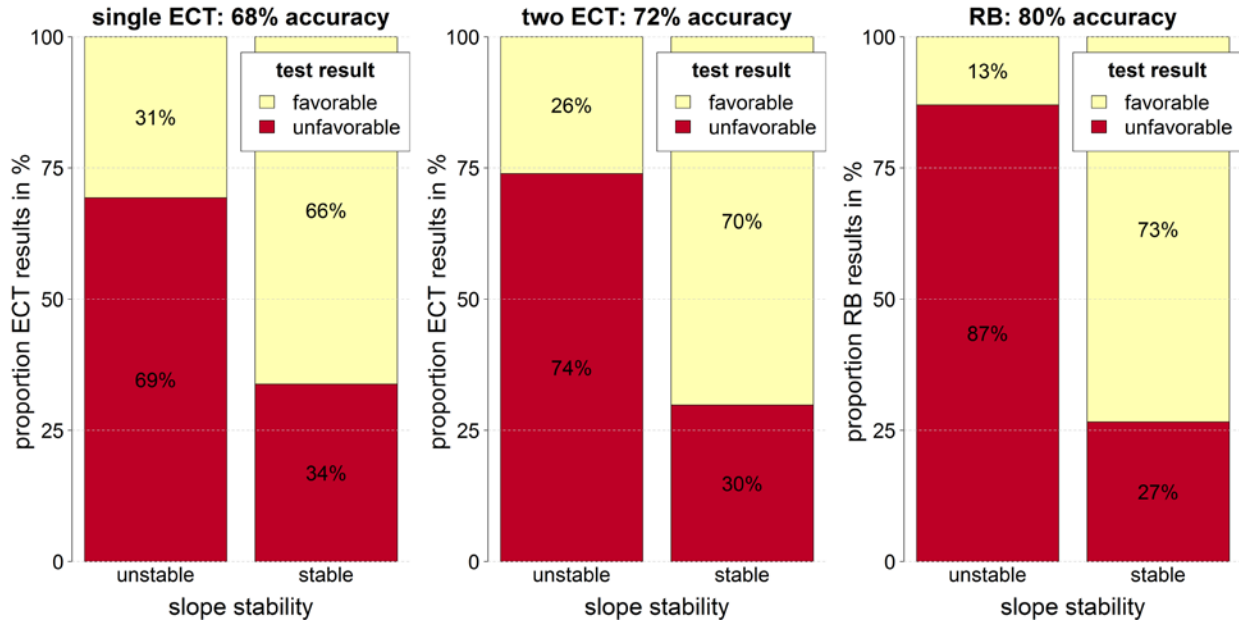


Fig. 1: Results of single ECT (left, N=442) and two ECTs with *favorable-unfavorable* test results (middle, N=177) and Rutschblock (right, N=140) in slopes rated as *unstable* and *stable*. As an example: for single ECT, 69% of the *unstable* slopes were correctly detected (test result *unfavorable*) and 66% of the *stable* slopes were correctly detected (test result *favorable*). The mean of these two values, 68%, is the unweighted average accuracy shown in Table 4.

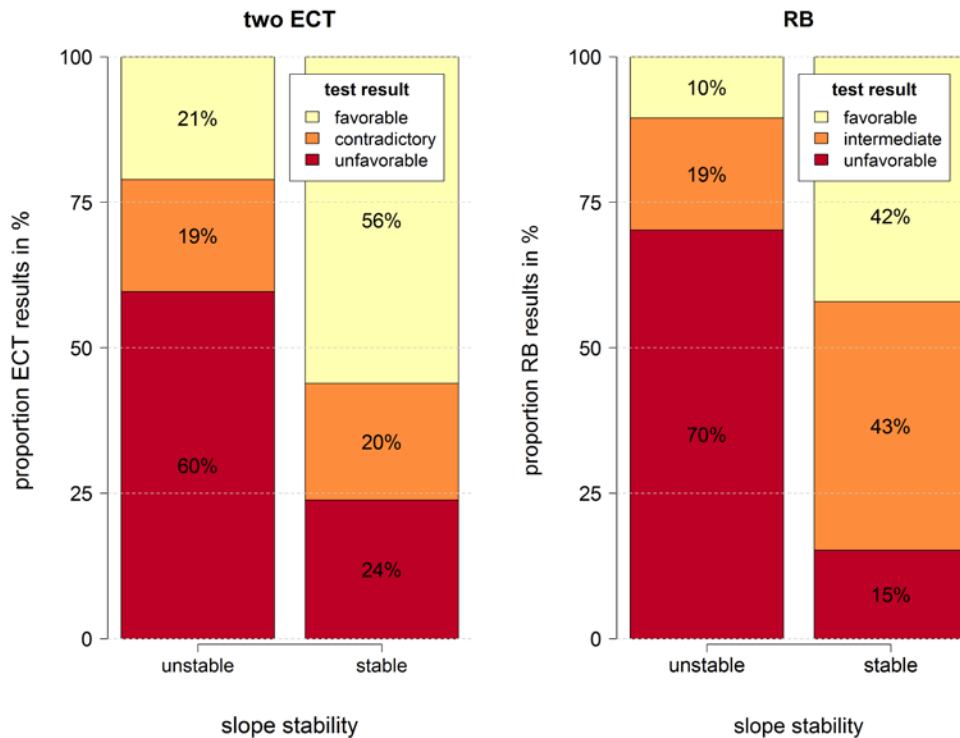


Fig. 2: Results of two ECT (N=221) and RB (N=221) including *contradictory* ECT and *intermediate* RB results for *stable* and *unstable* slopes.

4. DISCUSSION

4.1 *Performance of stability tests*

As shown in section 3.1, the repeatability of adjacent ECT test results was limited. This confirms Hendrikx et al.'s (2009) statement that a high correspondence between a single ECT test result (*unfavorable*, *favorable*) and slope stability rating (*unstable*, *stable*) cannot be achieved. In our dataset, this was mirrored in the low rate of slopes correctly detected as *stable* or *unstable* (68%) using the fracture propagation result of a single ECT. Performing a second test beside the first one, helped to increase the test accuracy slightly, when considering only two *favorable* or two *unfavorable* test results and slopes rated as *stable* or *unstable*.

About 20% of *contradictory* ECT results were observed. Their interpretation still remains unclear: a useful interpretation could neither be obtained by assigning *contradictory* ECT results to the slopes rated *stable* nor interpreting them as *unstable*. As was already pointed out by Winkler and Schweizer (2009), we suspect that *contradictory* ECT results may be a candidate for a *transitional* stability rating similar to *intermediate* RB results.

We achieved a lower test accuracy for single ECT than previous studies (Simenhois and Birkeland, 2009; Moner et al., 2008; Winkler and Schweizer, 2009). Possibly, this is due to using data collected by numerous different observers as part of operational slope stability evaluation rather than research. However, as we applied the same slope classification criteria to slopes tested with one or two tests, the test accuracy of both tests may be compared primarily within our study.

4.2 *Practical considerations*

A preferably high test accuracy is one criteria making a stability test valuable to the practitioner. While this favors the RB, other factors like the required time or the technical skill to perform a test also decide whether a test is useful or not (Schweizer and Jamieson, 2010). From our experience, it takes a lot more time and often two people to dig a RB, while an ECT (or two ECT) can easily and efficiently be performed by a single observer. Thus, to the avalanche professional we propose the following:

- If time permits, perform a RB as a stability test. It is the most accurate test and results are easy to observe, but it is time-consuming.

- If you don't have enough time or if you enter the selected slope for safety reason by yourself while your colleague waits at a safe spot, conduct two ECT rather than just one.

While we tested two ECT pairs performed in the same snowpit, Birkeland et al. (2010) suggested that avalanche professionals should not necessarily sample two stability tests adjacent to each other, but rather at some distance when searching for instabilities.

Even if done according to these recommendations, the interpretation of stability tests remains challenging. As has been remarked repeatedly, stability tests cannot be more than a single piece of the puzzle in the process of snow stability evaluation.

5. CONCLUSIONS AND OUTLOOK

We have explored a dataset of two adjacent ECTs, often together with a Rutschblock, in a diverse range of snowpack situations and performed by numerous different observers.

We have shown that in 21% of cases, the ECT result at the scale of the snowpit was not repeatable when performing a second test beside the first one. The RB performed better than the ECT in detecting *stable* and *unstable* slopes. However, the interpretation of ECT results may be somewhat improved by conducting two rather than just one test.

Future research will include snowpack information in combination with ECT results and explore test results in slopes with *transitional* snow stability, which were not considered in this analysis.

ACKNOWLEDGEMENTS

We thank the numerous SLF observers and employees, who have performed the stability tests and observed the snow-profiles. Without their effort, a study like ours would not have been possible.

We thank Lukas Dürr, Christine Pielmeier, Thomas Stucki and Jürg Schweizer for their valuable feedback.

REFERENCES

Bellaire, S., B. Jamieson, and J. Schweizer. 2010. When to dig? Thoughts on estimating slope stability. In Proceedings International Snow Science Workshop 2010, Lake Tahoe, CA: 424–30.

Birkeland, K., J. Hendrikx, and M. Clark. 2010. On optimal stability-test spacing for assessing snow avalanche conditions. *J. Glaciol.* 56 (199): 795-804

Darms, G., L. Dürr, and J. Schweizer. 2014. Field book for snow profile observations. WSL Institute for Snow; Avalanche Research SLF, Davos; Andermatt, Military mountaineering service centre, Andermatt.

Doswell, C., J. Davies, and D.L. Keller. 1990. On summary measures of skill in rare event forecasting based on contingency tables. *Weather Forecast* 5 (4): 576-85.

Föhn, P. 1987. The rutschblock as a practical tool for slope stability evaluation. *IAHS Publ.* 162: 223-28.

Hendrikx, J., K. Birkeland, and M. Clark. 2009. Assessing changes in the spatial variability of the snowpack fracture propagation propensity over time. *Cold Reg. Sci. Technol.* 56 (2-3): 152-60.

Moner, I., J. Gavalda, M. Bacardit, C. Garcia, and G. Marti. 2008. Application of field stability evaluation methods to the snow conditions of the Eastern Pyrenees. In *Proceedings International Snow Science Workshop 2008*, Whistler, Canada: 386-92.

R Core Team. 2016. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>. (last access: 18/08/2016)

Schweizer, J. 2002. The rutschblock test - procedure and application in Switzerland. *The Avalanche Review* 20 (5): 14-15.

Schweizer, J., and B. Jamieson. 2010. Snowpack tests for assessing snow-slope instability. *Annals of Glaciology* 51 (54): 187-94.

Simenhois, R., and K. Birkeland. 2006. The extended column test: A field test for fracture initiation and propagation. In *Proceedings International Snow Science Workshop 2006*, Telluride, CO: 79-85.

Simenhois, R., and K. Birkeland. 2009. The Extended Column Test: test effectiveness, spatial variability, and comparison with the propagation saw test. *Cold Reg. Sci. Technol.* 59 (2-3): 210-16.

Winkler, K., and J. Schweizer. 2009. Comparison of snow stability tests: Extended column test, rutschblock test and compression test. *Cold Reg. Sci. Technol.* 59 (2-3): 217-26.