WHAT'S THE SCORE
SCORING SNOWPITS FOR BETTER COMMUNICATION OF STABILITY ASSESSMENTS

Dave Richards[1], Karl Birkeland[2,3], Doug Chabot[4], Jim Earl[5]

[1]Avalanche Program, Alta Ski Area
[2]Birkeland Snow and Avalanche Scientific
[3]USDA Forest Service National Avalanche Center
[4]Gallatin National Forest Avalanche Center
[5]SnowPilot.org

ABSTRACT: Digging snowpits and utilizing those data for assessing stability is critically important for avalanche professionals who are managing and mitigating avalanche danger. However, quickly and efficiently *communicating* snowpit assessments to both professional peers and the public can be challenging. Sharing snowpit data and assessments is complicated, especially when sharing results with large numbers of people, such as over a radio at a ski area. To address this issue, we propose a snowpit scoring system – the *Grom Score* – that provides a rapid snapshot of the current stability.

Our system is based on three factors associated with snow stability: crack propagation propensity, ease of crack initiation, and snowpack structure. The *propagation propensity* score is based on whether or not an Extended Column Test (ECT) propagates, *ease of crack initiation* is scored by the number of ECT taps, and *snowpack structure* is scored on whether strong snow overlies weak snow and if a persistent weak layer is present. Each factor gets a score, we sum these scores (1-9), and higher values indicating increased stability.

To test the method's effectiveness, we first scored 100 snowpits randomly drawn from the SnowPilot database and compared pit scores to observer's stability ratings. Our initial encouraging results led us to expand our research by automating the scoring method and using it on 3,393 SnowPilot pits. This analysis showed pits with higher scores are more commonly rated as having "Very Good" or "Good" stability, while those with lower scores are typically associated with "Very Poor" or "Poor" ratings.

While most avalanche professionals will eventually want and need the nuanced information contained in a complete snow profile, our technique efficiently and quickly communicates basic stability information. Further, our scoring system provides both professionals and the public with a snapshot of stability.

1. INTRODUCTION

Often, practitioners are left in a quandary. After digging a snowpit they do not have a succinct and efficient method to communicate their findings to their peers. This makes it challenging to communicate relevant snow pit information to new riders, students, ski patrollers, and even the most experienced coworkers. Though we may understand snowpit data and what it tells us about the snow stability, communicating those findings quickly and succinctly is problematic. Our proposed method aims to simplify and improve communicating the most relevant data, though the nuance contained in snowpit profiles requires a closer look at the data. As a caveat, our scoring system applies only to dry slab avalanche conditions.

Though some practitioners previously viewed snowpits solely as a forecasting tool, the advent of newer testing methods gives us improved stability assessments and allows pits to be used as a now-casting tool. Sharaf and McCammon (2005) first looked at snowpit information through the lens of "Strength, Structure and Energy," rating each factor as Good, Fair or Poor. Many practitioners continue to use this method, and it is now taught in United States Level 1 curriculum. However, "Propagation Propensity" has replaced "Energy" since propagation can be indexed with modern stability tests while "Energy" has a clear physical meaning that is not captured in a typical snowpit (Figure 1).

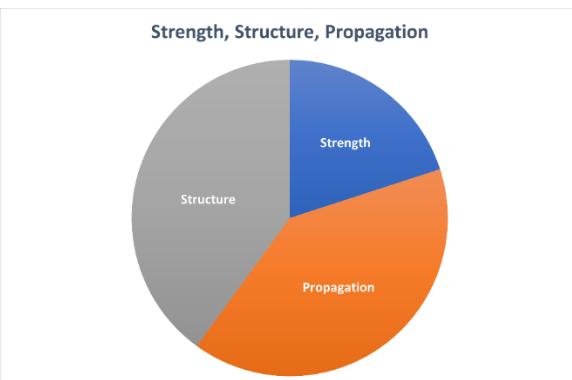

Figure 1: Snow strength, structure, and propagation propensity all contribute to assessing snow stability.

We prefer the nomenclature of Strength, Structure and Propagation, three things that everyone can see and interpret, and that require only a hand, a shovel and an Extended Column Test (ECT) cord. Further, these three factors can easily be given a score. The sum of these scores allows us to communicate snowpit findings in comparison to overall stability. When communicating snowpit information we need to do better than "so what?" or "looks good to me."

2. METHODS

2.1 Scoring pits

Researchers assess the utility of stability tests using methods that rate slopes as either stable or unstable (for example, Simenhois and Birkeland, 2009). However, the shortcoming of these techniques is that stability is not strictly binary. Techel et al. (2020) developed a more nuanced approach based solely on observed signs of instability and used that to better identify how well specific stability tests assessed stability.

Our proposed snowpit scoring technique sums up numerical scores for Strength, Structure, and Propagation propensity (Figure 2). Propagation propensity is indexed using Extended Column Test results, strength is scored with number of ECT taps, and structure is assessed by a simplified version of the *five lemons* (McCammon, and Schwiezer; 2002) called the PHD (persistent weak layer/hardness change/depth of the weak layer) factors of snowpack structure.

	Rating	Description	Score
Strength	Difficult	Tap score 21-30-ECTX	3
	Moderate	Tap score 11-20	2
	Easy	Tap score 0-10	1
Propagation	ECTX	ECT provided no failure	3
	ECTN	ECT provided no propagation	2
	ECTP*	ECT provided full propagation	1
	ECTPV*	ECT fails with full propagation on isolation	0
	* If ECTP structure ≤ 1		
	* If ECTPV structure = 0		
Structure	Good	Weak snow on top of strong snow, lacking a weak layer	3
	Fair	Strong snow on top of weak snow, greater than a meter deep and lacking a PWL	2
	Poor	Strong snow on top of weak snow, PWL is present	1
	Very Poor*	All lemons or PHD factors are present with PWL	0
	*If PWL than structure ≤ 1		
		Total Score =	

Figure 2: This table summarizes how pits are given a *Grom Score* based on scores for strength, propagation potential, and structure.

We feel that any scoring system should emphasize structure and propagation more heavily than strength, which can vary dramatically over short distances. In our scoring system we give additional weight to Structure and Propagation by scoring them on a scale of 0-3 rather than the 1-3 scale we use for strength, and by applying the following three rules:

***If PWL exists then Structure score ≤ 1**

***If ECTP then Structure ≤ 1**

***If ECTPV then Structure score = 0**

In our method, structure is scored using the three simple PHD factors. Is there a *persistent* weak layer (PWL)? Is there a hand *hardness* change of one step or greater? Finally, is the *depth* to the weak layer less than one meter? The PHD factors are the simplest method of scoring snowpack structure and are often taught in modern level one curriculum.

The two other scorable assessments – strength and propagation propensity – come from the Extended Column Test (ECT; Simenhois and Birkeland, 2006, 2009), a widely used and internationally accepted snow stability test. The test assesses the number of taps needed to initiate a weak layer crack, and it also provides an index of crack propagation propensity by observing whether that crack propagates across the entire column.

Discussions with many avalanche professionals about our proposed scoring system raised several questions. First, should propagation and structure be weighed more heavily than crack initiation? This is a valid criticism because if a crack cannot be initiated then there is no avalanche. However, crack initiation with the ECT is difficult when a weak layer is deeply buried, but that same weak layer may be more shallowly buried at other locations on that slope due to spatial variability. Second, several studies note that ECT propagation alone cannot perfectly discriminate between stable and unstable slopes (Simenhois and Birkeland, 2006; Moner et al., 2008; Simenhois and Birkeland, 2009; Winkler and Schwiezer, 2009; Techel et al., 2020). While this is true, research also shows that propagating ECT results are more often associated with unstable conditions (Techel et al., 2020). Therefore, when propagation in an ECT is present, our scoring method heavily weights it. Finally, some people argue structure trumps all. We agree that this is true when the weak layer is deeply buried. This is why our method scores the structure of any snowpack with a PWL as less than or equal to one.

At first glance, our method appears to score snowpits with more stable snow higher than those with less stable findings. For example, a 150 cm deep pit with a faceted weak layer at 90 cm and an ECTP 17 at that weak layer would score a four. This clearly indicates a situation with low stability. Another pit on a nearby slope might have the same faceted layer at the same depth, but with an ECTN 25. This improves our score to six, which is better but still not great. Finally, in a third pit we do not find the facets, but there is a non-persistent weakness 130 cm down where we get an ECTN 23. Our score for this pit is a seven, which is a slight improvement. Thus, in general terms, the higher a pit scores the better things seem to look. However, we wanted to more rigorously test the effectiveness of our scoring system.

2.2 Testing the scoring system

To assess whether or not our scoring system worked for differentiating between stable and unstable profiles, we utilized the SnowPilot database (Chabot et al., 2004; Snowpilot.org). We only included pits where an ECT score was reported, and the user gave the slope a stability rating. We first manually scored 100 snow pits randomly drawn from the database without prejudice to date or international location. Since these initial results were encouraging, we automated the scoring procedure, applied it to over 3,000 pits in the SnowPilot database, and graphed our results.

3. RESULTS AND DISCUSSION

Our analysis of 3,393 pits in the SnowPilot dataset shows promising relationships between user's stability assessments and numerical scores from our proposed method (Figure 3). With each increase in the pit score from 3 to 9, the proportion of pits that users rated as "Good" increases while the proportion of pits rated as "Poor" declines. This also holds generally for pits scoring a 1 or 2, but the number of pits in those categories are small so those results should be interpreted cautiously.

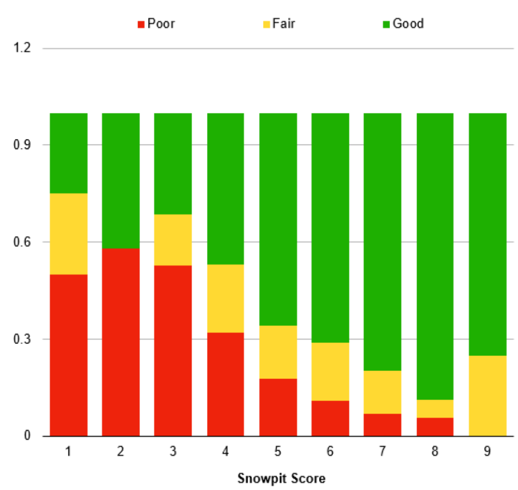

Figure 3: The proportion of stability ratings associated with each pit score. The number of pits with scores of 1, 2, and 9 are small, so those results should be interpreted cautiously. Note that lower scores are more commonly associated with poor stability and higher scores are more likely to be associated with good stability. Total N=3393.

A critically important takeaway from our results is the variability in stability assessments associated with each pit score. For example, even though most of the pits with a score of 3 are rated "Poor", about 30% of them are rated "Good". At the other end of the scale, although about 90% of the pits scoring an 8 are rated "Good", that still means about 10% of the pits are associated with stability ratings of "Poor" or "Fair".

We also note that our scoring system is relatively conservative, with more of a tendency to give a stable pit a low score (i.e., 30% of the pits with a score of 3 are rated "Good") than to give an unstable pit a high score (i.e., only about 5% of the pits with a score of 8 are rated "Poor" and no pits with a score of 9 are rated "Poor") (Figure 3). This is a desirable characteristic of our scoring system since it encourages conservative decision-making.

In a further analysis of 6,670 snowpits, pits were sorted into categories of stable (N=4472) versus unstable (N=2198) based on the parameters of Techel et. al (2020). This gives a 'base unstable rate' of .33 (Figure 4.). The Grom Score for snowpits aligns well with the Techel stability observations. At scores of four or less the number of pits rated unstable exceeds our base rate, at scores of 6 or more the unstable pits are less than the base rate, and at a score of five the number of pits rated unstable is roughly equal to our base rate. This gives us additional confidence that pits with lower scores are more likely to be associated with truly unstable snow, while pits with higher scores are more likely to be associated with stable snowpacks.

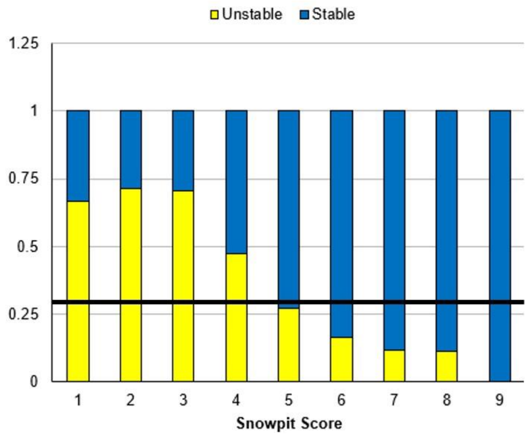

Figure 4: For this graph we classified 6670 pits in the SnowPilot database into "unstable" or "stable" bins based on the criteria used by Techel et al. (2020). From this we calculate that 33% of those pits are unstable, forming our "base rate" (shown by the black line). Here we see that at lower scores the number of unstable pits exceed our base rate, while with higher scores the number of unstable pits is less than our base rate.

Some of the variability in our results may be explainable; we noted a few cases where pits with low scores that users rated as having "Good" stability had unstable stability test results but relatively shallowly buried weak layers. However, other variations may be more challenging to explain. Ultimately, any pit scoring system – including ours – oversimplifies the more nuanced data contained in a complete snow profile. Thus, we strongly discourage using our pit scores to replace the important work of closely examining all the information contained in a pit profile. Instead, our method supplements the detailed information in snowpits by allowing avalanche professionals quickly communicate general information to other professionals or the public.

4. CONCLUSIONS

4.1 Facilitating communication with pit scores

The main goal of this work is to develop a method to quickly communicate snowpit information. Our pit scores can be easily communicated within an organization, to other avalanche professionals, or to the lay backcountry skier. While the *Grom Score* does not take the place of all the nuanced data contained in a snowpit, it does provide a quick and simple snapshot of stability based on snowpit information.

These scores can help untrained recreational users better understand pit results. Professionals can glean a great deal of useful information from a snowpit graph (see Figure 5), but these plots may confuse untrained backcountry skiers. Using the *Grom Score* we can communicate that this real-world pit scores a 3 and has poor stability. This pit scored zero for structure based on the presence of a PWL, depth, and a hardness change. It scored one for Propagation with an ECTP, and two for Strength with a tap of fourteen. The slope avalanched later in the week after this observation was made.

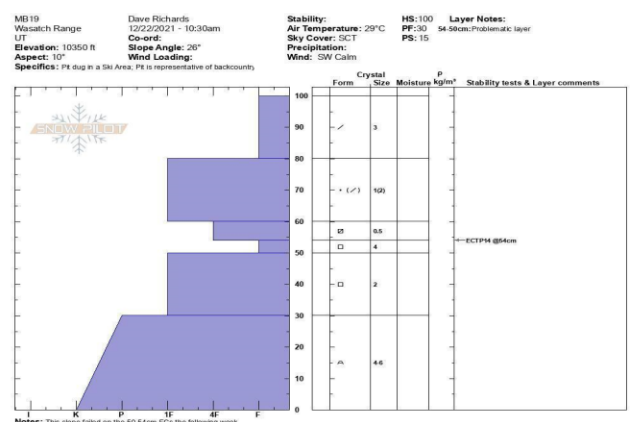

Figure 5: A snowpit with a *Grom Score* = 3. This slope avalanched later in the week.

Because the scoring method improves communication clarity, it may facilitate withingroup decision-making. When we rate stability from "very poor" to "very good", the words can be misconstrued and misunderstood. However, a numerical score is more easily understood and communicated. During the 2022/23 winter, the *Grom Score* was successfully used as a decision-making tool by numerous educators from both the University of Utah and the American Avalanche Institute.

One especially useful way to utilize pit scores is to map them on an aspect/elevation diagram (Figure 6). These graphs allow professionals and the public to condense all of the information contained in many snowpits into a simple, digestible diagram identifying current stability patterns. Using several of these diagrams over the course of a season provides a visual representation of changes in stability patterns. Avalanche educators and some heliski operations found this approach worked well during the 2022/23 winter.

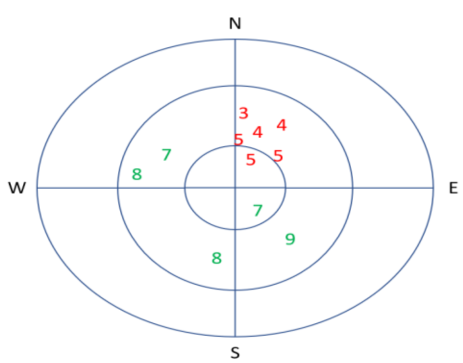

Figure 6: Example aspect/elevation diagram with pit scores. Such graphs may help users to better discern patterns of unstable snowpack in the backcountry or within an operation.

4.2 Summary

Avalanche professionals have discussed scoring snowpits for many years. Our method is a simple first step and a way to start the conversation with other professionals and researchers. That said, the initial feedback from a number of avalanche professionals is that this first step is a worthwhile effort toward improving the standardization and communication of pit results.

Analysis of a large database of thousands of snowpits dug worldwide show the *Grom Score* correlates reasonably well with the stability assessments of avalanche practitioners. Anecdotal evidence leads us to believe that a numerical score may be a better method of communication than our typical stability ratings.

For this study we only used the Extended Column Test results. However, some users do not apply this test. Perhaps a future iteration of our pit scoring method could utilize other stability tests such as Propagation Saw Tests.

Avalanche professionals work in an environment of increasingly complex data with even more complicated language to explain those data. We can simplify things by providing a snapshot of instability through our scoring method. Avalanche educator and helicopter ski guide Jim Conway summed things up when he told us: "The pit scoring system is simple to digest and use, and more importantly it allows quick concise communications in the field environment. The system still allows for more detailed traditional pit evaluation data to be shared when this is needed."

ACKNOWLEDGMENTS

The authors thank the following people for their invaluable contributions to this paper: Lynne Wolfe, Gabrielle Antonioli, Greg Gagne, Mark Staples, Drew Hardesty, Trent Meisenhiemer, John Tuckman, Liam Fitzgerald, Jonathan Morgan, Pete Groves, Jake Hutchinson, Jim Conway, and the staff of the Alta Ski Patrol.

REFERENCES

Chabot, D., M. Kahrl, K.W. Birkeland, and C. Anker. 2004. SnowPilot: A new school tool for collecting, graphing, and databasing snowpit and avalanche occurrence data with a PDA. *Proceedings of the 2004 International Snow Science Workshop, Jackson Hole, Wyoming.*

Moner, I., Gavaldà, J., Bacardit, M., Garcia, C., Mart, G. 2008. Application of Field Stability Evaluation Methods to the Snow Conditions of the Eastern Pyrenees. *Proceedings of the 2008 International Snow Science Workshop, Whistler, British Columbia.*

McCammon, I. and J. Schweizer. 2002. A field method for identifying structural weaknesses in the snowpack. *Proceedings of the 2002 International Snow Science Workshop, Penticton, British Columbia.*

Sharaf, D., and I.McCammon. 2005. Integrating Strength, Energy, and Structure into Stability Decisions, *The Avalanche Review* 23(3).

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

Simenhois, R. and K.W. Birkeland. 2009. The Extended Column Test: Test effectiveness, spatial variability, and comparison with the Propagation Saw Test. *Cold Regions Science and Technology* 59(2-3), 210-216.

Techel, F., K. Winkler, M. Walcher, A. van Herwijnen, and J. Schweizer. 2020. On snow stability interpretation of extended column test results, *Nat. Hazards Earth Syst. Sci.* 20, 1941–1953.