THE TAP-O-METER: HOW HARD DO PRACTITIONERS TAP AND SO WHAT?

Håvard B. Toft[1,2,*], Samuel V. Verplanck[3], Markus Landrø[1,2]

[1] *The Norwegian Water Resources and Energy Directorate, Oslo, Norway*
[2] *Center for Avalanche Research and Education, UiT - The Arctic University of Norway, Tromsø, Norway*
[3] *Department of Civil Engineering, Montana State University, Bozeman, MT, USA*

ABSTRACT: The hand-tap loading method is utilized in both Compression Tests (CTs) and Extended Column Test (ECTs). Discrepancy among the written standards of the United States, Switzerland, Norway, and Canada, as well as inherent subjectivity of the method's implementation motivate our study's objectives to characterize the impact force-time curves and quantify the variability between practitioners. We developed the "tap-o-meter", an instrument consisting of shovel blade mounted to a load cell which records the applied force during a hand tap at a rate of 50 kHz. This sampling rate enables precise measurement of both peak force and loading rate, the two metrics chosen to characterize the applied loads. We collected data from 286 practitioners, including avalanche forecasters and mountain guides in Scandinavia, Central Europe, and North America to generate a data set of 8522 individual taps. For each loading step (wrist, elbow and shoulder), the inner quartile peak forces are distinctly different, as are the loading rates. However, there is significant overlap across the range of measurements, with instances of participants applying more force in wrist taps compared to the shoulder taps of others. This overlap challenges the reliability and reproducibility of ECT results, potentially leading to dangerous interpretations in avalanche decision-making, forecasting and risk assessments. The data collection process was paired with a short survey which included weight, height, gender, country of residency, and avalanche climate. The information in these explanatory variables did not explain the bulk of the variance. The only explanatory variable that is significantly correlated with tap force across all multivariate regression models is gender. Our results provide an answer to the question of "How hard do we tap?" but not necessarily "How hard should we tap?". We believe these data and insights will enable discussion among the tests' creators, the scientific community, and the practitioner community to update thresholds, guidelines, and test interpretation.

Keywords: stability tests; impact force; hand taps; tap-o-meter; standardization; decision-making

1. INTRODUCTION

Stability tests simulate different parts of the avalanche release process which aids in decision making. Furthermore, mathematical modeling of stability tests has the potential to connect stability tests to slope wide avalanche mechanics. The extended column test (ECT) is frequently used, yet there are different interpretations of tests results and no mathematical model. A key component of this test is the hand tap loading method. Numerous written standards exist for the hand tap loading method (e.g. Greene et al. (2022); Campbell et al. (2016); Dürr and Darms (2016); NVE (2022)), and they all have discrepancies such as whether:

- To use fingertips, knuckles, open hand or fist at different loading steps.
- Letting the hand fall with its own weight or applying downward force until a descriptive pain tolerance level is reached.

The objective of our work is to characterize the force-time curves of the hand tap loading method. See Toft et al. (2024a) for the full version of this study.

2. METHODS

2.1 The measurement device: "tap-o-meter"

To measure the force from hand taps, a device dubbed the "tap-o-meter" was made (Figure 1). A total of three devices were built to enable data collection in different parts of the world in a similar time frame. Each "tap-o-meter" has the following components:

- A shovel blade which acts as the loaded surface.
- A load cell to transduce the tapping force into an electric signal.
- Oscilloscope with a voltage amplifier to measure the signal.
- 30 x 30 x 0.6 cm stainless steel base to provide a sturdy foundation.

It is important to determine an appropriate sampling rate for the tap-o-meter. We are most interested

---

[*] Corresponding author address:
Håvard Boutera Toft, The Norwegian Water Resources and Energy Directorate
Middelthuns gate 29, 0368, Oslo, Norway;
tel: +47 45482195;
email: htla@nve.no

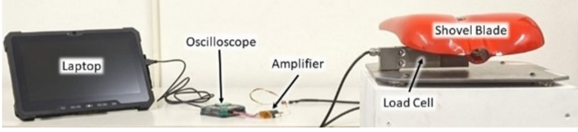

Figure 1: The “tap-o-meter” consists of a metal base with the load cell and shovel blade attached above. The load cell is connected to the oscilloscope through the custom-built 201x amplifier.

in the peak force and loading rate leading up to it. Preliminary testing showed that the rise time from hand taps can happen as quickly as a few milliseconds. Conservatively assuming this rise occurs over 1 millisecond, a sampling rate of 50 kHz leads to 50 samples in this critical measurement period. This number was deemed sufficient for our purposes and within the capabilities of the measurement system.

See Toft et al. (2024) for a more thorough description of the tap-o-meter and its components.

2.2 The data collection procedure

Data collection was conducted at events in Norway, Switzerland, Austria, USA, and Canada. In Norway, data was collected from avalanche forecasters and mountain guides. In Switzerland, data was collected at the European Avalanche Warning Service (EAWS) general assembly. Canadian and Austrian events only included avalanche forecasters. Events in the USA contained a mix of avalanche workshop participants and avalanche forecasters. A total of 286 individuals (232 males and 54 females) contributed to the study. A detailed table of the number of samples, event, and date can be found in Toft et al. (2024a). We did not provide any specific instructions on how to conduct the ECT other than that we asked participants to tap as they would do in the field. We provided a wide range of gloves with different thicknesses, but it was up to the participants themselves to select which glove, or whether to use a glove at all.

We made the setups as similar as possible by using three identical “tap-o-meter” devices. All “tap-o-meters” were firmly attached to a wooden CT (30 x 30 x 85 cm) or ECT (30 x 90 x 85 cm) column (Figure 1). By using a fixed height, we acquired data with a consistent sampling method but are not able to adjust for changes in simulated snowpack thickness. Furthermore, participants were given the choice to use different types of gloves depending on their preferences. The intent was that all participants should be able to conduct the test like they would do in the field. However, we left the shovel handle off as early tests during the development showed that even gentle touches are picked up by the sensitive load cell.

For each participant, we asked them to fill out a survey where they noted their country of residency, avalanche climate, height, weight and gender. The information from the survey was collected to answer the following research questions:

- Does height, weight, and/or gender affect tapping force?
- Does height, weight, and/or gender affect tapping force?
- Does height, weight, and/or gender affect tapping force?

2.3 Data processing

The raw data from the tap-o-meter is a time-series table of voltage readings (Figure 2). The tap-o-meter was calibrated using known weights ranging from 50 to 300 newtons which allowed us to convert the voltage readings to newtons. We used a peak finding algorithm to identify individual taps as a first pass, followed by a more manual process to ensure accurate results. After the taps are identified, two metrics are pulled from each one: peak force (newtons, N) and loading rate (N/s). More details on the data processing procedure can be found in Toft et al. (2024a).

2.4 Idealization of taps as Gaussian functions

The taps are idealized as Gaussian functions to visualize the data and to provide a steppingstone for future mathematical modeling efforts of the ECT. Both peak force, F_{peak}, and loading rate, r, are used to idealize the observed taps. The following relation was used to generate impact force, F, as a function of time, t:

$$F(t) \approx F_{\text{peak}} \cdot e^{-\frac{1}{2}\left(\frac{3r(t-t_{\text{peak}})}{F_{\text{peak}}}\right)^2} \tag{1}$$

where t_{peak} is the time at which the peak of the Gaussian occurs (0.025 s, in this case).

3. RESULTS

3.1 Peak force and loading rate

The data set consists of 2,837 wrist taps, 2,839 elbow taps, and 2,846 shoulder taps across 286 individuals. Outliers are excluded using 1.5 times the interquartile range (IQR) method, which is a widely recognized and accepted standard in statistical analysis (Tukey, 1977).

In Table 1 and 2, we provide some descriptive statistics of peak force and loading rate. The median peak force approximately doubles from one

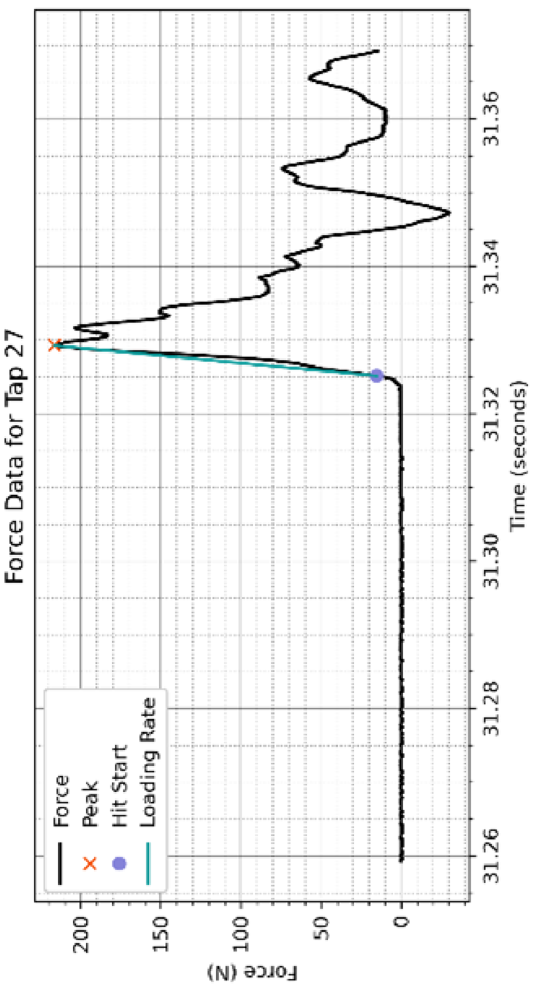

Figure 2: An example of the data processing procedure implemented on a shoulder tap. This procedure acquires two metrics for each tap: peak force (N) and loading rate (N/s).

loading step to the next at 79 N, 185 N and 373 N respectively. The standard deviation is also roughly half of the mean peak force for each loading step, showing that the variability in loading increases proportionally with increasing peak force. The loading rate, and its standard deviation, increases with each load step. The loading rate is positively correlated with peak force ($R^2 = 0.70$).

We observed different mean and median values for each loading step, and if we consider the interquartile range, which represents the data between the 25th and 75th percentile, there is nearly no overlap between loading steps. Doing a one-way ANOVA, we get a p-value lower than 0.01, indicating that the three loading steps are statistically different from each other, mirroring the findings of Sedon (2021) and Griesser et al. (2023).

Table 1: Descriptive statistics of peak force (N)

	Peak Force (N)		
	Wrist	Elbow	Shoulder
Mean	79	185	373
Std. Dev.	39	82	172
Min	8	34	45
25th percentile	50	123	239
Median	73	173	343
75th percentile	101	237	481
Max	190	426	893

Table 2: Descriptive statistics of loading rate (N/s)

	Loading Rate (N/s)		
	Wrist	Elbow	Shoulder
Mean	8,819	28,836	66,088
Std. Dev.	6,745	17,362	41,951
Min	118	149	2,316
25th percentile	3,449	15,107	37,128
Median	6,842	25,068	61,553
75th percentile	12,763	39,830	90,676
Max	30,145	81,619	195,812

In Figure 3, the distribution of peak forces across different tap numbers is graphically represented for three tapping levels. While the median forces across each loading step remain relatively consistent, there is a large spread across all loading steps. Collectively, this figure emphasizes the inherent differences in peak forces across the three tapping levels and underscores the variability present within each level across different tap numbers.

3.2 Gaussian function idealization

Using the median metrics along with their 25th and 75th percentiles (Tables 1 and 2), the idealized force curves are shown in Figure 4.

From this figure, we see that a tap from the shoulder is generally a sharper pulse (i.e. shorter duration, higher peak force) than a wrist tap, with an elbow tap in between the two. Using this idealization, we estimate the median loading duration to be 21 ms for the wrist, 14 ms for the elbow, and 11 ms for the shoulder.

4. DISCUSSION

Using the data from the "tap-o-meter", we can provide insight into the impact forces of hand taps and the variability between participants. We believe the quantification of the magnitudes and variabilities associated with hand-tap loading will assist with our understanding and interpretation of the ECT and CT.

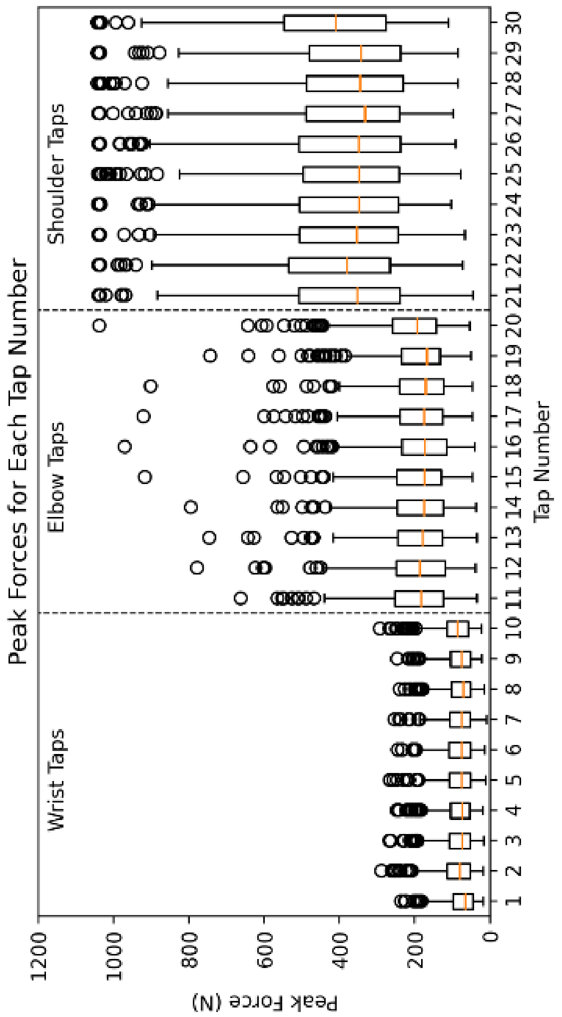

Figure 3: A visualization of the magnitude and variability in peak impact force from the 286 participants from tap 1 to 30. A box plot for each tap number dis-plays the minimum, first quartile, median, third quartile, and maximum values. Outliers are shown using circular symbols. The load cell reaches saturation at 1,000 N, a threshold which was reached in one elbow tap and 75 shoulder taps.

4.1 Body characteristics, gender and region

Our main finding from the survey data is that only gender has a statistically significant relationship with peak force. Body features (weight and height) are also correlated with peak tap force, but when included in a multivariate analysis with gender, they disappear.

Given the variations in observational guidelines for the ECT, we hypothesized that measuring differences among participants from the Alps, Scandinavia, and North America would be feasible. Despite this expectation, we observed no regional variations in peak tapping force. The lack of significant findings might be attributed to our limited predictive capability from the small sample size in a statistical context (n=286), or that there are no differences to be found.

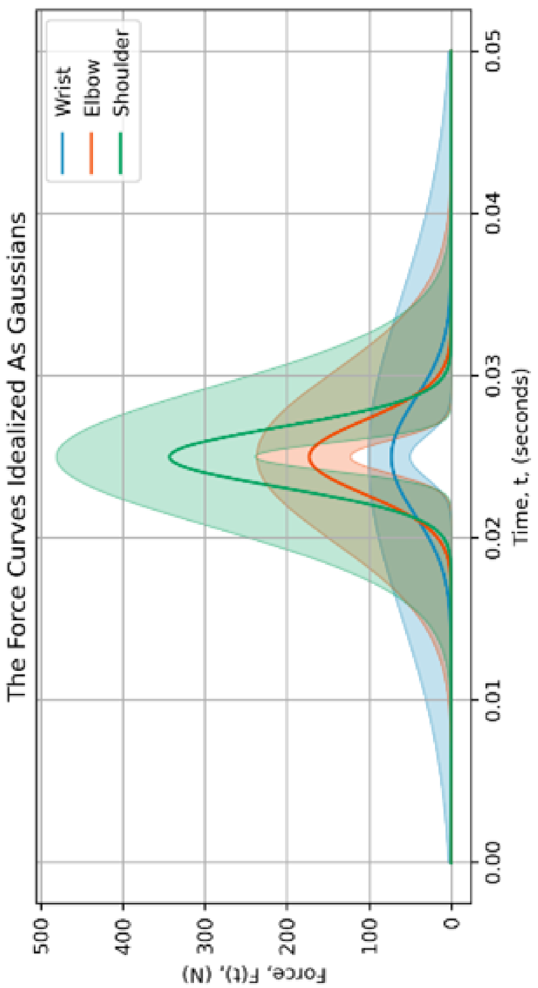

Figure 4: An idealization of the taps as Gaussian functions. The center lines are from the median metrics and the shading is generated from the 25th and 75th percentiles.

4.2 Variability in tapping force – implications for stability interpretations

It is widely agreed that whether a crack propagates across the entire column or not is the key discriminator between unstable and stable slopes (Techel et al. (2020)). However, both Winkler and Schweizer (2009) and Techel et al. (2020) show that the number of taps provides additional information, allowing a more refined distinction between results related to stable and unstable conditions. Techel et al. (2020) found the optimal threshold between ECTP20 and ECTP22, which aligns with the ECTP21 threshold suggested by Winkler and Schweizer (2009). Moving away from a binary classification came at the cost of introducing intermediate stability classes (Techel et al. (2020)). These new intermediate stability class definitions rely heavily on the tap number when failure occurs. Variability in the applied force-time curves likely leads to variability in test results, particularly regarding the number of taps required to induce weak layer failure. It is important to emphasize that no tests offer a definitive "go/no go" result. With accuracies of around 80%, these tests are not reliable enough to be the main factor in our slope scale decision-making (Birkeland et al. (2023)).

We found the three loading steps to have statistically different IQRs; this aligns with the results from Griesser et al. (2023), which highlight this as a positive outcome and that the CT and ECT hand-tap procedure is somewhat reliable. Despite the statistical differences in each loading step, we question the application of average results to individual cases. The main difference in our argument lies in relying solely on mean statistics to develop tapping norms used by individuals. For example, we found that 18% and 26% of elbow taps have a peak force value that falls within the tapping norms for wrist and shoulder taps, respectively. This implies that 44% (18% + 26%) of elbow taps would be misclassified as taps hinging from the wrist or shoulder. Assuming peak applied force influences test results, then this misclassification of loading steps will lead to a misclassification of test results. Because stability tests results aid in an individual's decision-making process, a misclassification of test results could lead to dangerous consequences in real-world applications.

4.3 Implications for avalanche practitioners

Given the variability in tapping demonstrated in this study, we propose two considerations to improve the ECT standards. The two ideas outlined below are intended to be a foundation for further discussion in the broader avalanche community.

4.3.1 Reduce tapping variability through the use of training and/or tools.

The large variability in impact force between individual participants highlights the need for standardization. This could be done by creating a better definition of how the test should be conducted in terms of technique and tapping force. When interpreting the descriptive definitions from each loading step, it is impossible to infer which impact forces should be used as a baseline for each loading step. For example, the Norwegian description (Norwegian Water Resources and Energy Directorate, 2022) using the arm's weight would depend on the weight of each participant's arm. Furthermore, using Canada as an example, there is no description of how hard each tap should be other than that it should not hurt at shoulder level (Campbell et al. (2016)). However, this would depend on the participant's pain tolerance, snow properties (dampening) and the participant's glove thickness.

The community will need to agree on what the ideal impact force-time curves are. The impact forces presented in this paper could be used as a baseline for future clarifications if a "wisdom of crowds" impact force definition is employed (see Surowiecki (2005) for an introduction to the concept of "wisdom of crowds"). An alternative to the "wisdom of the crowds" is a selection of experts could choose to define the appropriate windows and thresholds.

With these windows defined, a training device could be developed that measures the impact force and reports back to the participants whether they are within the correct window at each hand loading step. If a training device is considered to be the best solution to reduce interpersonal variability, we believe this study provides sufficient information to build such a training device. Such devices already exist for CPR training and provides real-time measured feedback on compression rate (cpm), depth (mm), release (g), compressions count, and inactivity time during CPR, while also enabling responders to self-evaluate their performance with event statistics on the spot (Laerdal (2023)).

Another solution could be to develop a tool that ensures consistent impact force, like the stuffblock test (Johnson and Birkeland (1998)). The test involves filling a nylon sack with 4.5 kg of snow and dropping it in increments of 10 cm. However, this type of loading has its challenges. The peak force and loading rate are coupled and depend on the object's mass, the drop height, and the materials that are in contact during impact. Not only mass and height would need to be recommended, but also materials and possible use of cushion-like material to recreate both peak force and loading rate of hand taps. Verplanck and Adams (2024) attempted to match the impact curves of hand taps using an acetal mass, foam cushion, and aluminum plate. However, they attempted to match their own hand taps, not the averages presented in our study.

4.3.2 Revisiting the stability interpretation of CT and ECT

Our second proposition comes from the implication of defining predictor thresholds based on impact forces from a large database of ECTs. The concern is that the large variability in hand-tap loading makes these average-based thresholds relatively weak. The thresholds make sense when analyzing large amounts of data (e.g. in the context of avalanche forecasting) but not when applying the average results to individual cases. We should therefore evaluate whether the importance of the number of taps outweighs the risk of misinterpreting the test result.

One thought example could be whether it is appropriate to interpret ECTP20 (intermediate stability) compared to ECTP24 (unstable) in individual cases (Winkler and Schweizer (2009)), given the large discrepancies in impact force. There is also precedent for adopting a more straightforward approach in interpreting ECT results at the expense of leaving potentially relevant information out, as when shear quality and fracture characteristics were removed from the ECT (Ron Simenhois and Greene (2018)). In this approach, we would consider the test result to be unstable if crack propagation occurs, and stable otherwise. When using the more simple, binary approach, the impact force becomes less important, and the large variation is less of a problem.

4.4 Future work

While our study has made strides in accurately observing the force-time curves from hand taps, there are still areas that require further exploration. For instance, tap force measurements greater than 490 N may not be as accurate force measurements below 490 N because 0-490 N is the recommended load cell range (see Toft et al. (2024a)). Also, our calibration assumes the load cell responds similarly to dynamic loads as static loads and eccentric loads as centered loads. These potential inaccuracies in the measurement technique likely contribute to the range and variability of force measured in this study. Future studies should therefore include a load cell with a higher range (e.g. 2000 N), load cells designed for impacts (e.g. piezo-resistive), and a fixture to ensure centered loading. By doing so, we can enhance the precision, accuracy, and reliability of our measurements, leading to more robust and accurate findings. Despite these potential measurement inaccuracies, our study utilized a sampling rate (50 kHz) appropriate for capturing the entirety of the impact curve. This is an improvement over similar studies that used a sampling rate of 100 Hz (Griesser et al. (2023)) and 105 Hz (Thumlert and Jamieson (2015)). Sedon (2021) do not provide any sampling rate for their study.

Furthermore, systematic notes about the tapping technique would also be of interest. A qualitative remark is that many of the participants infrequently use their fingertips on wrist taps as in the North American standards (Greene et al. (2022); Campbell et al. (2016)). There was also a large variability in impact forces because of different techniques such as using the weight of the arm versus a shoulder tap so hard that it hurts the hand.

During the data collection, we asked participants if they regularly conduct CTs or ECTs for work, recreation or both. For future studies, a more effective approach might involve quantifying the frequency of CTs or ECTs performed by each participant per season. This method could provide a more nuanced understanding of the relationship between the quantitative experience and tapping consistency.

5. CONCLUSION

In this study, we developed a device that can accurately measure force-time curves from the hand-tap loading method. The dataset collected is the largest one to date (286 participants, 8522 taps), including data from Scandinavia, the Alps, and North America. From these data, we quantified peak force and loading rate for each tap, both of which increased for each loading step (i.e. wrist, elbow, shoulder). There is nearly no overlap in peak force from the 25th to 75th percentile between loading steps. Yet there is significant overlap in the outer quartiles with examples of some wrist taps with as high of peak force as others' shoulder taps. Assuming peak applied force influences stability test results, then this misclassification of loading steps will lead to a misclassification of stability test results.

Using the observed peak forces and loading rates, the force-time curves are idealized as Gaussian functions. This idealization provides a convenient steppingstone for future mathematical modeling efforts of stability tests like the Compression Test and Extended Column Test. We investigated whether the differences in weight, height, gender, and/or geographical region influence peak force using multivariate statistical models. Overall, these variables explain very little of the variance in peak tap force, with over 90% of the variance attributed to factors other than height, weight, gender, and geographical region. Our results indicate that gender is the only statistically significant explanatory variable, with women's peak force being approximately 20% less than men's peak force.

Our results provide an answer to the question of "How hard do avalanche practitioners tap?" but not necessarily "How hard should avalanche practitioners tap?". We recommend that our data be used to facilitate discussions related to updating guidelines for the hand-tap loading method, possibly of including thresholds of peak force and loading rate for each loading step. Given the variability in tap-ping demonstrated in this study, we propose two considerations to improved standards: (1) reduce tapping variability through the use of training and/or tools and (2) evaluate whether the importance of the number of taps outweighs the risk of misinterpret-ing the stability test results.

6. DATA AVAILABILITY

The data needed to replicate the study is available in our Open Science Framework repository (Toft et al. (2024b)).

ACKNOWLEDGEMENT

We would like to acknowledge Knut Møen for his technical contributions to the development of the "tap-o-meter" and for his creative input in naming the device. Furthermore, Andrea Mannberg for her statistical expertise, Christoph Mitterer and Scott Savage for assistance in data collection – with additional thanks to Scott for facilitating us while working on this in Idaho. Jordy Hendrikx for connecting the authors, a collaboration born out of the realization that we were doing similar work. Thank you to all the study participants as well.

References

Birkeland, K. W., van Herwijnen, A., Techel, F., Bair, E. H., Reuter, B., Simenhois, R., Jamieson, B., Marienthal, A., Chabot, D., and Schweizer, J.: Comparing stability tests and understanding their limitations, in: Proceedings of the 2023 International Snow Science Workshop, Bend, OR, 2023.

Campbell, C., McClung, D., Jamieson, B., Sayer, B., Whelan, R., Floyer, J., and Garvin, S.: Observation guidelines and recording standards for weather, snowpack and avalanches, Canadian Avalanche Association, 6th edition edn., 2016.

Dürr, L. and Darms, G.: SLF-Beobachterhandbuch (Observation guidelines), WSL Institute for Snow and Avalanche Research SLF, Davos, URL https://www. slf. ch/fileadmin/user_upload/WSL/Publikationen/Sonderformate/pdf/SLF-Beobachterhandbuch. pdf, 2016.

Greene, E., Birkeland, K., Elder, K., McCammon, I., Staples, M., Sharaf, D., Trautman, S., and Wagner, W.: Snow, Weather and Avalanches: Observation Guidelines for Avalanche Programs in the United States, American Avalanche Association, 4th edition edn., 2022.

Griesser, S., Pielmeier, C., Toft, H. B., and Reiweger, I.: Stress measurements in the weak layer during snow stability tests, Annals of Glaciology, pp. 1–7, doi:10.1017/aog.2023.49, 2023.

Johnson, R. and Birkeland, K.: Effectively using and interpreting stability tests, in: Proceedings International Snow Science Workshop, Sunriver, Oregon, USA, 27 September-1 October 1998, pp. 562–565, 1998.

Laerdal: CPRmeter 2 User Guide, Laerdal, URL `https://cdn.laerdal.com/downloads/f6537/cprmeter_2_user_guide_en`, 2023.

NVE: Felthåndbok for Snø og Skredobservasjoner, Norwegian Water Resources and Energy Directorate, 3rd edition edn., 2022.

Ron Simenhois, Doug Chabot, K. B. and Greene, E.: Shear Quality or Fracture Character with an Extended Column Test – No Longer in SWAG or SnowPilot, American Avalanche Association, p. 1, 2018.

Sedon, M.: Evaluating forces for extended column tests and compression tests, Canadian Avalanche Journal, 127, 39–41, 2021.

Surowiecki, J.: The wisdom of crowds, Anchor, 2005.

Techel, F., Winkler, K., Walcher, M., van Herwijnen, A., and Schweizer, J.: On snow stability interpretation of extended column test results, Natural Hazards and Earth System Sciences, 20, 1941–1953, doi:10.5194/nhess-20-1941-2020, 2020.

Thumlert, S. and Jamieson, B.: Stress measurements from common snow slope stability tests, Cold Regions Science and Technology, 110, 38–46, doi:10.1016/j.coldregions.2014.11.005, 2015.

Toft, H. B., Verplanck, S. V., and Landrø, M.: How hard do avalanche practitioners tap during snow stability tests?, Natural Hazards and Earth System Sciences, 24, 2757–2772, doi:10.5194/nhess-24-2757-2024, 2024a.

Toft, H. B., Verplanck, S. V., and Landrø, M.: Tap-o-meter data, URL `https://doi.org/10.17605/OSF.IO/BV5PM`, dataset, 2024b.

Verplanck, S. V. and Adams, E. E.: Dynamic models for impact-initiated stress waves through snow columns, Journal of Glaciology, pp. 1–15, doi:10.1017/jog.2024.26, 2024.

Winkler, K. and Schweizer, J.: Comparison of snow stability tests: Extended column test, rutschblock test and compression test, Cold Regions Science and Technology, 59, 217–226, doi:10.1016/j.coldregions.2009.05.003, 2009.