How hard do avalanche practitioners~~we~~ tap during snow stability tests?

Håvard B. Toft[1,2], Samuel V. Verplanck[3] and 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 Mechanical Engineering, Montana State University, Bozeman, MT, 59717

Correspondence to: Håvard B. Toft, Norwegian Water Resources and Energy Directorate, Oslo, Norway; tel: +47 454 82 195; email: htla@nve.no

Abstract. This study examines the impact force applied from hand taps during Extended Column Tests (ECT), a common method of assessing snow stability. The hand-tap loading method has inconsistencies across the United States, Canadian, Swiss, and Norwegian written standards, as well as inherent subjectivity. We developed a device, the "tap-o-meter", to measure the force-time curves during these taps and collected data from 286 practitioners, including avalanche forecasters and mountain guides in Scandinavia, Central Europe, and North America. Peak forces and loading rates are the metrics chosen to quantitatively compare the data. The mean, median, and inner quartile peak forces are distinctly different for each loading step (wrist, ~~elbow~~elbow, and shoulder), as are the loading rates. However, there is significant overlap across the range of measurements and examples of participants with higher force wrist taps than other participants' shoulder taps. This overlap challenges the reliability and reproducibility of ECT results, potentially leading to dangerous interpretations in avalanche decision-making, forecasting and risk assessments. Our results provide an answer to the question of "How hard *do* avalanche practitioners~~we~~ tap?" but not necessarily "How hard *should* avalanche practitioners~~we~~ tap?". T~~We believe t~~hese data and insights ~~will~~ are intended to facilitate ~~enable~~ discussion among the tests' creators, the scientific community, and the practitioner community to update thresholds, guidelines, and test interpretation~~Therefore, we recommend updating the standards for the ECT. We propose two viable paths for future action: (1) define a target impact force-time curve for each tap level and develop tools and training to minimize variability in tapping force (2) assess the significance of the information derived from the number of taps. If deemed not highly valuable, we should consider reverting to a simpler binary interpretation that focuses exclusively on crack propagation~~.

1 Introduction

Snowpack instability describes the propensity for a slope to avalanche and has been modeled to include the mechanics concepts of failure initiation and crack propagation as key components of the avalanche release process (Reuter & Schweizer, 2018). Stability~~These~~ tests help gather crucial information on weak layer identification, failure initiation, and crack propagation. In our~~this~~ paper, we will often use~~use the colloquial~~ terms "snowpack~~lope~~ stability" and "stability tests", rather than "snowpack~~lope~~ instability" and "instability tests", due to their widespread usage in the avalanche practitioner community. Determining snowpack stability is a core concept in avalanche forecasting and backcountry decision-making, yet it is a challenging measure to quantify. ~~In the context of this introductiondiscussion, snowpack stability describes the propensity of a snow-covered slope to avalanche (Reuter & Schweizer, 2018). It is inversely related to the probability of avalanche release~~

and, therefore, a core concept in avalanche forecasting and backcountry decision-making. In backcountry travel, the decision process ultimately ends with a go or no-go decision based on an assessment of avalanche likelihood, avalanche size, and potential consequences. Snowpack stability evaluation is essential in assessing avalanche likelihood in such a context. To aidsimplify this complex decision-making process, snow stability tests have been invented. They provide a structured analytical approach, particularly valuable when direct signs of ininstabilitystability, like recent avalanches, shooting cracks, or whumpfs, are absent. weak layer identification,,n

In an avalanche forecasting setting, snow stability is divided into four classes: very poor (natural / very easy to trigger), poor (easy to trigger (e.g., a single skier), fair (difficult to trigger (e.g., explosives), and good (stable conditions). To determine the avalanche danger level for a region or forecast area at a point in time, forecasters in Europe use a matrix combining snowpack stability, snowpack stability frequency distribution, and avalanche size (Müller et al., 2023). Combined, snow stability and frequency are an expression of likelihood. The frequency distribution of snowpack stability is described using four classes: many, some, a few, and none or nearly none. Using the European Avalanche Warning Service (EAWS) matrix, a combination of "very-easy-to-trigger," size 3-avalanches in "many" slopes leads to a danger level 4-high, whereas "difficult to trigger," size 3-avalanches in "many" slopes equals a danger level 3-considerable. This shows how an increase in stability effects the danger level, frequency being the same.

In situations with poor snowpack stability, nature provides apparent signs such as recent avalanches, shooting cracks, and "whumpfs". These signs are commonly referred to as Class I factors (ininstabilitystability factors) in a three-class division (LaChapelle, 1980; D. McClung & Schaerer, 2006). TUnfortunately, the associated uncertainty increases for each class. The more stable the snowpack, the greater the load it can support before it fails. The inInstability can be less evident in these situations, and more indirect factors (class II -snowpack factors and class III meteorological factors) must be evaluated. However, d. Hence, stability tests (class II) can be of great importance in avalanche forecasting and provide highly valuable information to the backcountry traveler. To determine the avalanche danger level for a region or forecast area at a point in time, forecasters in Europe use a matrix combining snowpack stability, snowpack stability frequency distribution, and avalanche size (Müller et al., 2023). Previous studies have shown/ indicated that there is a difference in tapping force between those who perform tapping tests, which has a direct effect on test results. Combined with the associated uncertainty with class II factors in general, and the variability of the hand tap loading, the results should be used with this understanding in mind.

One of the first documented field snow tests is the shovel shear test developed by Faarlund and Kellermann in 1974 (originally known as the Norwegermethode; Kellermann, 1990). Although the role of compressive stress in weak layer failure was in discussion at the time (Perla & LaChapelle, 1970), weak layer shear strength - measured with a shear frame – was a typical metric for slope stability, and the shovel shear test provided a convenient field method of obtaining similar information.

In the late 1980s, Föhn (1987) quantified the Rutschblock (RB) test into the seven known levels today. In the 1990s, the compression test (CT) became popular (Clarkson, 1993; Jamieson & Johnston, 1996). Both ~~T~~the CT and RB involve loading the snow surface, transmitting stress through the slab, and the possibility of~~which may lead to~~ weak layer failure. A~~The key~~ distinction between these tests lies in their load application method~~triggering mechanisms~~: the CT~~ test~~ ~~is initiated by~~utilizes hand-taps, while the RB test requires the load of a person on skis.

The propensity for an initiated crack to propagate became a popular concept as a collapse-based, crack-propagation model (Heierli et al., 2008) had conflicting results with a shear-based, crack-propagation model (D. McClung, 1979). In line with this discussion, the propagation saw test (PST) (Gauthier & Jamieson, 2008, 2006) and extended column test (ECT) (Simenhois & Birkeland, 2006) were developed as field tests to assess propagation propensity.

~~Currently, T~~The ECT is a frequently used test by avalanche practitioners and recreationists. The test has been validated in different geographies and avalanche climates such as continental and intercontinental climates of the United States (Birkeland & Simenhois, 2008; Hendrikx & Birkeland, 2008; Simenhois & Birkeland, 2009), the Swiss Alps (Techel et al., 2020; Winkler & Schweizer, 2009), the Spanish Pyrenees (Moner et al., 2008) and New Zealand (Hendrikx & Birkeland, 2008; Simenhois & Birkeland, 2006).

The four stability tests described above measure different types of information in the snowpack using different triggering mechanisms, set-ups, and dimensions. Relevant types of information are whether the test can (1) identify weak layers in combination with slabs, (2) measure failure initiation, and (3) measure crack propagation. We have summarized the properties of each test in Table 1 with inspiration from Birkeland et al. (2023).

Table 1: Different types of information that can be extracted from the four different stability tests (modified from Schweizer and Jamieson, 2010; Birkeland et al. 2023).

Test	**Identifying weak layer below slab**	**Measures failure initiation**	**Measures crack propagation**	**Triggering mechanism**	**Dimensions (width, upslope)**
RB	Yes	Yes	Yes	Weight of a human	2 m x 1.5 m
CT	Yes	Yes	Partly	Hand-tap	30 cm x 30 cm
ECT	Yes	Yes	Yes	Hand-tap	90 cm x 30 cm
PST	No	Partly	Yes	Cutting with saw	30 cm x 100 cm[1]

[1] or the weak layer depth, whatever is greater.

As is evident in Table 1, ~~S~~stability tests are meant to simulate portions of~~ the~~ avalanche release process~~s on a small scale~~. T~~he key t~~o connect~~ing~~ stability tests with slope-wide avalanche mechanics, ~~is~~ a mathematical model of the stability test is needed. To date, most of this modeling has been done with the PST (Benedetti et al., 2019; McClung & Borstad, 2012; Van Herwijnen et al., 2016; Weißgraeber & Rosendahl, 2023). A key component of the ECT is the hand-tap loading which creates a boundary condition for a mathematical model of the ECT. ~~Although~~Creating this model is out of our scope~~ this paper does not attempt to create a mathematical model of the ECT~~, however, characterizing the impact curves are an important step towards modeling the ECT.

~~1.~~ To conduct an ECT, the~~a~~ hand-tap loading method is implemented that was originally developed for the CT. There are subtle differences in the current guidelines for these hand-taps.~~ ~~ The American Avalanche Association (2022) defines the most recent US standard as follows. This is similar to the Canadian standard (Canadian Avalanche Association, 2016), which has expanded the definition by including the text marked with *italics*.~~The most recent US standard, by American Avalanche Association (2022), defines it as:~~

~~1. “Tap 10 times with fingertips, moving hand from wrist.”~~

~~2. “Tap 10 times with the fingertips or knuckles moving forearm from the elbow.”~~

~~3. “Hit the shovel blade moving the arm from the shoulder 10 times with open hand or fist”.~~

~~and the Canadian standard (Canadian Avalanche Association, 2016):~~

~~2.~~1. “Tap 10 times with fingertips, moving hand from wrist.”

~~3.~~2. “Tap 10 times with the fingertips or knuckles moving forearm from the elbow. *While moderate taps should be harder than easy taps, they should not be as hard as one can reasonably tap with the knuckles*”.

~~4.~~3. “Hit the shovel blade moving the arm from the shoulder 10 times with open hand or fist. *If the moderate taps were too hard, the operator will often try to hit the shovel with even more force for the hard taps - and may hurt his or her hand*”.

In other countries~~places of the world~~, the instructions~~ may~~ vary as well~~even more. as f~~ For example, in Switzerland~~ where~~ the instructions are simply described using a single sentence: «*The blade of the avalanche shovel is placed on the block on one side and successively loaded with 10 hits each from the wrist (1-10), the elbow (11-20) and the shoulder (21-30).* » (Dürr and Darms, 2016). There are further discrepancies if we look at the

~~And the~~ Norwegian standard (Norwegian Water Resources and Energy Directorate, 2022).

“For every sequence of 10 taps, the load is increased as follows:

1. Let the hand fall with its own weight, lifted from the wrist.
2. Let the hand and forearm fall with their own weight, lifted from the elbow.

3. Let the entire arm fall with its own weight, using a fist, lifted from the shoulder.”

If a failure in the snowpack is detected during any of the taps, the specific tap number along with the depth of the weak layer is recorded for further investigation. For example, if a failure propagates at the 21st tap at a depth of 40 cm, it would be noted as 'ECTP21@40cm. The interpretation of ECT test results remains an open discussion. Originally, a binary interpretation of test results was suggested, referred to as ECT_{orig} in this paper. Specifically, if a fracture initiates but does not propagate (ECTN), then the test result is considered stable. In contrast, if a fracture propagates across the extended column (ECTP, or ECTPV if during isolation), then the test result is considered unstable. If no fracture is initiated within the 30 taps, the outcome is neither stable nor unstable, and should therefore be regarded as inconclusive.The test result is ECTX if no fracture occurs, resulting in an inconclusive result..

Another classification was suggested by Winkler and Schweizer in 2009 (ECT_{w09}), using three classes divided by the number of taps needed to initiate a fracture with or without propagation:

- ECTP ≤21 – low stability
- ECTP >21 – intermediate stability
- ECTN or ECTX – high stability

Recent work by Techel et al. (2020) (ECT_{t20}) suggests using four classes and applying the established labels for snow stability: poor, fair and good (ei.ge. American Avalanche Association, 2022):

- ECTP ≤13 – poor
- ECTP >13 to ECTP ≤22 – poor to fair
- ECTP >22 or ECTN ≤10 – fair
- ECTN >10 or ECTX – good

The variability in tapping force has been a known limitation for the CT and ECT interpretation (American Avalanche Association, 2022; Schweizer & Jamieson, 2010; Techel et al., 2020). Birkeland and Johnson (1999) attempted to remedy this limitation by developing the stuffblock test. The test uses a nylon sack filled with ~4.5 kg (10 pounds) of snow which is dropped on a CT or ECT column with 10 cm increments until a failure initiationresult is reached.

There have been some previous studies to measure the applied force of hand tapping, as well as quantify the stress-state within the snow during these loads. Logan (2006) made measurements of hand taps during a conference to learn more about timing, impact force and technique, but the results were never published. Thumlert and Jamieson (2015) impacted the snow with both a drop hammer and hand taps and measured the resulting stress within the snow. Our study expands on~~is very similar to~~ the ~~ones from the descriptive study from~~ work of Sedon (2021) and Griesser et al. (2023). ~~Both studies have made important contributions to the measurements of hand-tap loading, but they also have their inherent limitations. Both studies have a small sample from one defined geographical area (n=69 and n=62 respectively).~~ Each of these studies measured tap force by avalanche practitioners (n = 69 and n=62, respectively) in an indoor setting. Furthermore, Griesser et al. (2023)~~they~~ investigate the effects of body characteristics such as weight and height. Their analyses consist of bivariate tests, i.e., testing if people who are heavier tap harder, and if people who are taller tap harder. A limitation ~~problem~~ of~~with~~ this approach is that, since height and weight are typically correlated, the tests do not reveal which of the two factors that are more important, or if height (weight) affect tap ~~hardnes~~forces at a given weight (~~(~~height). Regarding sampling rate, a critical aspect of accurately measuring dynamic loads, ~~More importantly,~~ Sedon (2021) does not specify their~~s sampling rate~~ and Griesser et al. (2023) use a sampling rate of 100 Hz (one measurement every 10 ms). ~~The typical loading duration of a single hand-tap is between 20 and 50 ms, which means that Griesser et al. (2023) only had 2-5 measurements within this critical period. The low sampling rate makes it unlikely that the measurement device can accurately capture the peak force, which is shorter than a couple of milliseconds.Griesser et al. (2023) measured the impact force from 62 participants from the European Avalanche Warning Services, as well as a field-based study where they measured stress within the snowpack.~~

The objective of our work is to develop an improved measurement device that can accurately characterize the impact curves of~~the~~ hand-tap loading and investigate the interpersonal variability between participants from different geographical regions. W~~Additionally, w~~e plan to use multivariate regression to investigate whether body characteristics, snow climate, and gender influence the impact force from hand taps. Furthermore, we intend to not only measure the peak force, but also the loading rate, a metric not included in the previous studies by Sedon (2021~~0~~) and Griesser et al. (2023). It has been well-established that snow's response depends on the loading rate (Shapiro et al., 1997), a quantity shown to both influence stress wave transmission through snow slabs (Verplanck and Adams, 2024) as well as failure of weak layers such as depth hoar, facets, and surface hoar (Reiweger et al. 2015). Thus, peak force alone is not enough information to accurately understand and predict snow's response dynamic loads. Determining how ~~this applied force interacts with~~ snow responds to the applied force from a hand-tap is outside of our scope, however, a quantified understanding of how hard practitioners tap~~are loading the snow~~ will~~ould~~ aid in the process of updating standards for test execution and interpretation.~~test result interpretation. In addition, this study will aid in the mathematical modeling of stability tests.~~

~~An additional novelty in our study is the measurement of loading rate. It has been established that snow's response depends on the loading rate (Shapiro et al., 1997). Thus, peak force alone is not enough information to accurately model snow's response dynamic loads. Snow subject to shorter duration impacts (~1 ms) have been observed to attenuate at shallower depths than show subject to longer duration impacts (~10 ms) (Verplanck and Adams, 2023). Loading rate has also been shown to influence~~

~~crack initiation in weak layers of snow (Reiweger et al. 2015). Furthermore, observing both peak force and loading rate will enable the specification of a boundary condition for future modeling efforts of the CT and ECT.~~

~~By modeling stability tests, researchers can gain insight into the avalanche release process and quantify key parameters such as the effective elastic modulus of the slab and weak layer fracture energy (van Herwijnen et al., 2016). Modeling efforts have focused on the PST rather than the CT or ECT (Benedetti et al., 2019; McClung & Borstad, 2012; Weißgraeber & Rosendahl, 2023). One reason is the crack is initiated by cutting through the weak layer with a snow saw during the PST, a more measurable mechanism than the impact-initiated crack of the CT/ECT. Although our study does not attempt to model the CT/ECT, characterizing the force-time curves of hand taps is a step towards a more scientific understanding of these tests.~~

2 Methods

2.1 The device: "tap-o-meter"

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

1. A shovel blade which acts as the loaded surface.
2. A load cell to transduce the tapping force into an electric signal.
3. Oscilloscope with a voltage amplifier to measure the signal.
4. 30 x 30 x 0.6 cm stainless steel base to provide a sturdy foundation.

2.1.1 Load cell

A single, cantilever-style load cell from Load Cell Central (GCB3-SS-M-50KG) was used to measure the tapping force. The recommended capacity of the load cell is 490 N, with an ultimate overload rating of 1,470 N. The full-scale output (FSO) of the load cell is 2 mV/V and refers to the maximum output signal that the load cell can produce for its rated capacity.

2.1.2 Oscilloscope and voltage amplifier

An oscilloscope (Digilent Analog Discovery II) was used to measure the impact force. The oscilloscope provides a 5-volt input to the load cell, which yields a maximum output signal of 10 mV with the FSO from the load cell. The minimum change in voltage that can be measured by the oscilloscope is 0.2 mV. To increase the measurement resolution, a linear voltage amplifier was added between the load cell and the oscilloscope. The amplifier was custom built using an AD8429 amplifier from Analog Devices. The amplification, or gain (G), is controlled by an external two-pin resistor (R_{ext}), using the following equation:

$$G = 1 + \frac{6000\ ohm}{R_{ext}}, \tag{1}$$

In our study, we used a 30-ohm resistor, resulting in a 201x amplification of the output signal from the load cell. Using this setup, the oscilloscope is theoretically able to measure 10,050 steps between 0-490 Newtons, or 30,150 loading steps between 0-1,470 N. The device was calibrated statically by using a set of known weights~~masses~~ ranging from ~50 to 300~~0~~ N~~kg~~ (Appendix-1), resulting in ~~and~~ a linear regression with ~~(~~R^2 = 0.999998~~)~~.

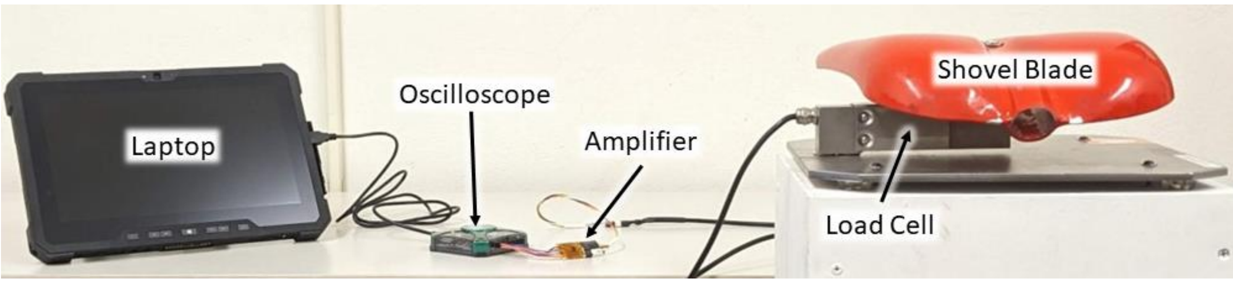


Figure 1: The "tap-o-meter" consists of a metal base with the load cell and shovel blade attached above. The load cell is connected to the oscilloscope through the custom-built 201x amplifier.

To determine an appropriate sampling rate, knowledge of the ~~impact~~ signal is critical. We are most interested in the peak force and loading rate leading up to it. Preliminary testing showed that this rise time is fastest for the shoulder taps and 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. A number deemed sufficient for our purposes and within the capabilities of the measurement system.

The "tap-o-meter" was initially developed using parts in stock at the Norwegian Water Resources and Energy Directorate (NVE). Early testing suggested that a ~500 N load cell which NVE had in stock would be capable of accurately recording the impact force from taps. Based on data collected prior to those showcased in this paper, it became evident that the impact forces from some participants plateaued around 600 N on their shoulder taps. This level surpassed the recommended operating range of the load cell but stayed within the ultimate overload capacity (~1,500 N). We pinpointed the problem to the amplifier, which was reaching its saturation point.

We considered the amplifier properties to avoid two potential issues. Setting it too high would mean losing detail in measuring light wrist taps due an increased background noise. On the other hand, setting it too low would make it impossible to measure the strongest impact forces.

To address this, we developed a new adjustable amplifier that we tuned to a range from 5 to 1,000 N. This calibration aimed to balance the ability to detect high-impact forces while maintaining a low background noise for measuring the force of lighter taps. The defined range stayed safely below the load cell's ultimate overload threshold of 1,225 N. Despite the new adjustment with the amplifier's upper limit set to 1,000 N, saturation still occurred in rare instances: once during elbow-level taps (representing 0.03% of such taps) and 75 times for shoulder-level taps (2.63% of such taps).

2.2 Data collection process

Data collection was conducted at events in Norway, Switzerland, Austria, USAUSA, and Canada. In Norway, data was collected from avalanche forecasters and mountain guides. In Switzerland, data was collected at the 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 Appendix-2x. 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.

Table x: A description of each event, date and number of samples gathered.

Event	**Date**	**Samples**
Montana State University Ski and Avalanche Workshop	26.10.2022	25
Friends of the Gallatin National Forest Avalanche Center	15.11.2022	9
Ski saw	18.11.2022	30
Sawtooth Avalanche Center	08.03.2023	26
Instructor TrainingPro dev	05.04.2023	17
outhwest Montana Ski Patrol Snow Science Day **(Sam?)**Forecasters at Parks Canada	24.02.2023	4
Mountain Guides Meeting, Innsbruck #1Colorado Avalanche Information Center	02.03.2023	5
Mountain Guides Meeting, Innsbruck #2Chugach National Forest Avalanche Information Center	13.03.2023	3
European Avalanche Warning Services General Assembly	15.06.2022	62
Galtuer	15.12.2022	15
Grillhof	30.11.2022	17
Norwegian Avalanche Observer Workshop	08.11.2022	46

Gallatin National Forest Avalanche Center Professional Development Workshop	12.11.2022	27
(Sam?)UIAGM General Assembly Norway		

We made the setup 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 (Fig. 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 with our sensitive load cell .. No samples were collected with the shovel handle on.We did this because the user may grab the shovel handle and torque the load cell, leading to an inaccurate measurement.

2.2.1 Survey

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:

1. Does height, weight, and/or gender affect tapping force?
2. Do people tap differently across avalanche climates?
3. Are there regional differences between Scandinavia, Alps and North America?

2.3 Data processing

The raw voltage data are processed using python to identify the individual taps. After the taps are identified, two metrics are pulled from each one: maximum force (newtons, N) and loading rate (N/s). To gain a quantitative understanding of the collected data, peak force and loading rate were the chosen metrics. Other quantities such as impact duration, rise time, and stress were considered but not chosen. Impact duration was not used because the measurements frequently contained long, oscillatory tails that are artifacts of the load cell rebounding and vibrating – a phenomenon expected to be less present during an actual field test. Rise time iwas calculated as an intermediary step to loading rate. However, loading rate was chosen because snow's response has been shown to depend on its rate of deformation (Shapiro et al., 1997, Reiweger et al., 2015, Verplanck and Adams, 2024). Lastly, our measurements are presented as forces (N) rather than stresses (kPa) because presenting it as a stress would rely on an assumption of cross-sectional area.

The recorded time and voltage are imported as NumPy arrays (Harris et al., 2020). The voltage values are zeroed by subtracting the entire array's mean from each data point. Then, voltage is converted to newtons by scaling according to the calibration.

Scipy’s (Virtanen et al., 2020) peak finding algorithm, scipy.signal.find_peaks, is implemented to determine when the taps occur by comparing neighboring values. The peak finding algorithm is driven with two parameters: a 25 N minimum peak magnitude and 0.4 seconds minimum time between peaks. These criteria are chosen by iteratively trying different values and viewing the results. This peak finding method is used as a first pass through the data and is later refined with a more manual process. See Figure 2 for an example of tap data with the peaks algorithmically identified.

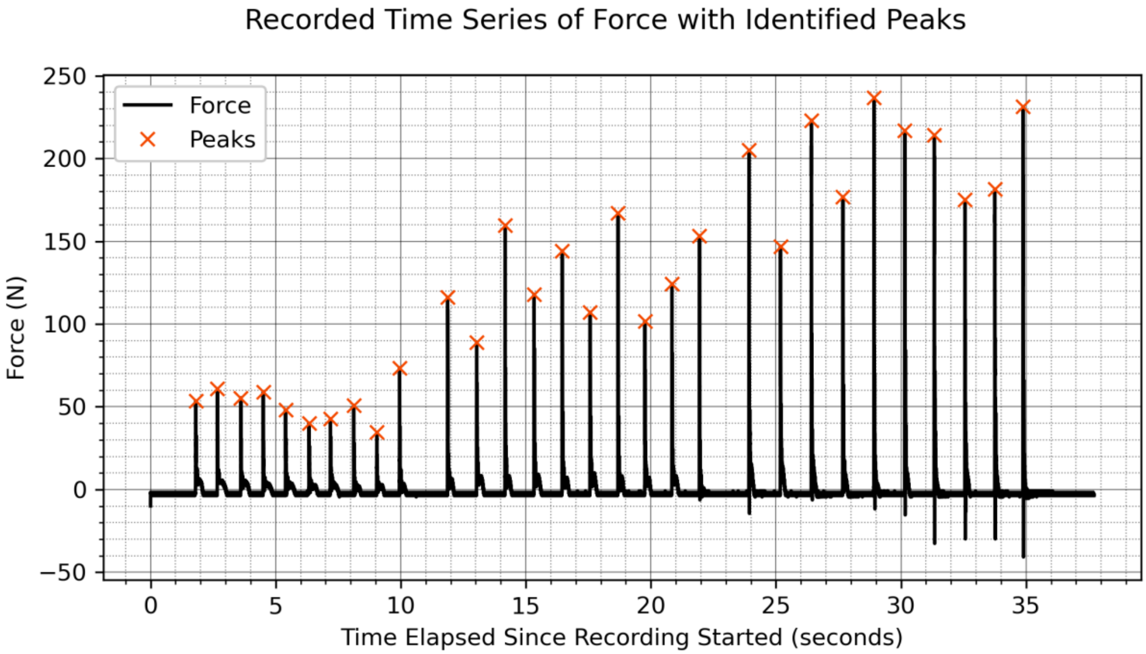


Figure 2: ~~The result~~An example of identifying taps using SciPy’s peak finding algorithm with 25 N minimum peak magnitude and a minimum of 0.4 seconds between peaks. Using these parameters, the algorithm correctly identified all peaks as it did in 262/286 cases. Manual adjustments to the algorithm’s parameters were used in the remaining 24 cases to identify peaks.

After the peaks are found the individual taps are defined as 70 ms prior to and 40 ms after the peak. These values are chosen to allow for enough time surrounding the peak to determine tap metrics. Each tap array is then re-zeroed by subtracting the mean of the first 0.2 ms of that specific tap. This re-zeroing process is implemented because subtle shifts in the baseline recording are occasionally apparent, particularly during the taps hinging from the wrist if the tapper kept contact with the shovel blade throughout these taps. The two metrics, maximum force and loading rate, are ascertained from each tap array. Maximum force, F_{peak}, is simply the maximum value in the re-zeroed array. The loading rate, r, is defined as a linear interpolation, Eq. (2), between the maximum force, F_{peak}, and a threshold value greater than typical noise, λ. In our measurements, a λ of 15 N was deemed appropriate. ~~when the force curve crosses a 15 N threshold, signifying the start of impact. A threshold chosen because it is greater than typical noise.~~ The difference in force is divided by the rise time, Δt, to determine the loading rate. The rise time is the difference in time between the peak force and the initial threshold crossingstart of the tap.

$$r\,Loading\ rate = \frac{(F_{peak}-\lambda)}{\Delta t}, \qquad (2)$$

After this automated process is applied to all 286 tap recordings, a manual quality control process is done. This process entails viewing the taps for each recording (Fig. 3), flagging misidentified taps, and classifying which taps are hinging from the wrist, elbow and shoulder. This manual process determined that 262/286 recordings were correctly processed with the first-pass algorithm. The remaining 24 recordings were reprocessed by changing the parameters for SciPy's peak finding algorithm. The changes to peak-finding parameters involved reducing the time between peaks or minimum magnitude until all the clear taps are identified. In some cases, the metrics were not calculated accurately because there was a spike of noise that was close enough in time to the tap signal. In these cases, the individual taps were not included in the analyzed data set. After this second processing step, the data set is ready for analysis.

To gain a quantitative understanding of the collected data, peak force and loading rate were the chosen metrics. Other quantities such as impact duration, rise time, and stress were considered but not chosen. Impact duration was not used because the measurements frequently contained long, oscillatory tails that are an artifact of the load cell rebounding and vibrating — a phenomenon expected to be less present during an actual field test. Rise time was calculated as an intermediary step to loading rate. However, loading rate was chosen because snow's response has been shown to depend on its rate of deformation (Shapiro et al., 1997). Lastly, our measurements are presented as forces (N) rather than stresses (kPa) because presenting it as a stress would rely on an assumption of cross-sectional area.

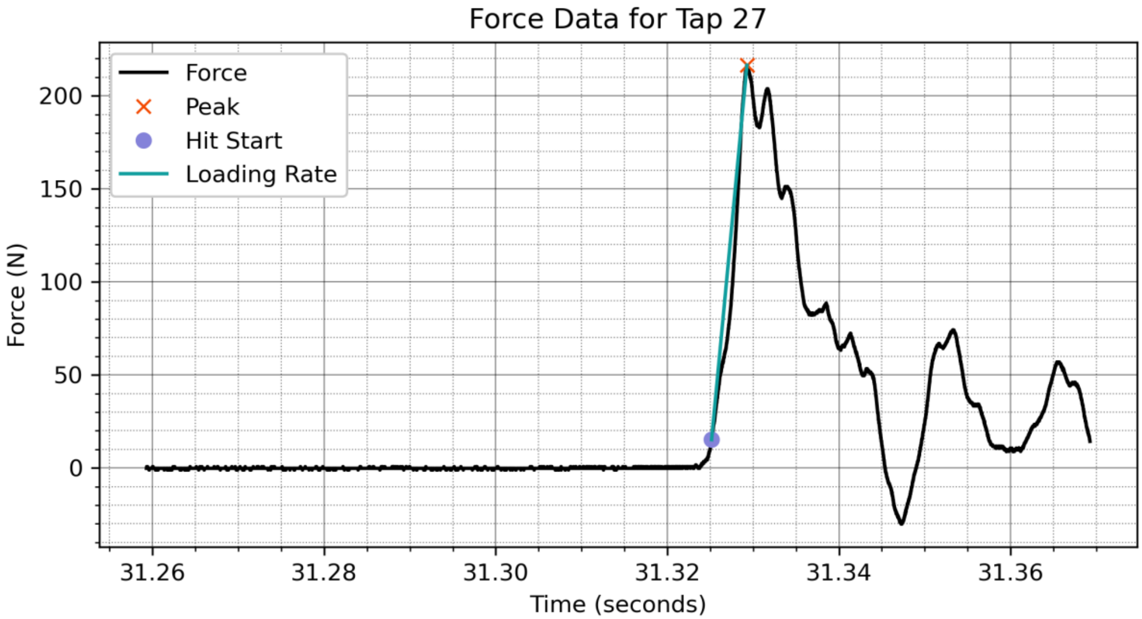


Figure 3: An example of the data processing procedure implemented on a shoulder tap. This procedure acquires two metrics for each tap: peak force (N) and loading rate (N/s).

2.4 Statistical analysis

We performed a set of ordinary least squares (OLS) regression models to understand the underlying factors influencing hand-tap loading. More specifically, we tested height, weight, gender, and geographic region on impact force during tapping tests. The peak force was the dependent variable in these models. A one-way ANOVA was conducted to assess whether wrist, elbow and shoulder taps were statistically different. All analyses were considered statistically significant at p-values below 0.05.

3. Results

3.1 Trends and variability by individual tappers

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). For peak force, we excluded 119~~32~~ taps (4.2%) for wrist, 93~~106~~ taps (3.3%) for elbow and 123~~0~~ taps (4.3%) for shoulder as outliers. Saturation occurred in rare instances due to a limitation with the amplifier in the "tap-o-meter":

it happened once (~0.0%) during elbow taps and 75 times (2.6%) during shoulder-level taps -(Table 21). We provide more details onMore on this in section 4.1.1.

Table 21: Number of taps, outliers and saturation taps for peak force and loading rate.

	Peak Force			Loading Rate		
Tap Level	Wrist	Elbow	Shoulder	Wrist	Elbow	Shoulder
No. of taps	2,837	2,839	2,846	2,837	2,839	2,846
No. of outlier taps	119 (4.2%)132	93 (3.3%)106	123 (4.3%)0	149 (5.2%)5	108 (3.8%)16	205 (7.2%)4
No. of saturation taps	0 (0.0%)	1 (0.0%)	75 (2.6%)	0 (0.0%)	0 (0.0%)	0 (0.0%)

In Table 32, we provide some descriptive statistics of peak force and loading rate (outliers removed using 1.5 times IQR). The median peak force approximately doubles from one 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 tap levelloading step (wrist, elbow, and shoulder) all tap, showing that the variability in loading increases proportionally with increasing peak force. The loading rate, and its standard deviation, increases with each load step (i.e. tap levelloading step). The loading rate is positively correlated with peak force ($R^2 = 0.70$).For peak force, the average wrist tap force recorded is 85 N with a standard deviation of 47 N. At elbow taps, the average force is higher at 196 N, with a standard deviation of 104 N. The shoulder taps are the powerful having an average peak force of 400 N, with a standard deviation of 211 N. The maximum peak force for the wrist, elbow, and shoulder levels are 190 N, 426 N, and 893 N respectively (Table 2).

The loading rates are presented for the same three tap levels. The wrist taps, have an average loading rate of 11,263 N/s with a standard deviation of 34,563 N/s. The elbow taps, show an average loading rate of 31,752 N/s with a standard deviation of 23,179 N/s. The shoulder level has the highest average loading rates at 86,179 N/s with a standard deviation of 79,202 N/s. The maximal loading rates for the wrist, elbow, and shoulder levels are 30,145 N/s, 81,619 N/s, and 195,811 N/s respectively (Table 2).

Table 32: Descriptive statistics of peak force and loading rate (outliers removed using 1.5 * IQR).

	Peak Force (N)			Loading Rate (N/s)		
Tap Level	Wrist	Elbow	Shoulder	Wrist	Elbow	Shoulder
Mean	7985	18596	373400	8,81911,263	28,83631,752	6686,088179
Std.	3947	82104	172211	6,74534,563	17,36223,179	41,95179,202

Min	8	34	45	118	149	2,316
25th percentile	501	12355	23944	3,449572	15,107550	37,1288,771
Median	734	17355	34352	6,8427,379	25,068982	61,5535,560
75th percentile	1017	23745	481505	12,7634,238	39,83042,016	90,676101,673
Max	190	426	893	30,145	81,619	195,8121

~~3.1 Trends and variability by individual tappers~~

In Figure 4, the distribution of peak forces across different tap numbers is graphically represented for three tapping levels: wrist, elbow, and shoulder. For each tap number, a boxplot illustrates the interquartile range, with the median force denoted by an orange horizontal line. Individual outliers, shown as circles, showcase the spread of peak forces. While the median forces across each ~~tap level~~loading step remain relatively consistent, there is a large spread across all ~~tap level~~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. Another method of visualizing the statistical spread of the data is shown in Appendix-3 with a confusion matrix.

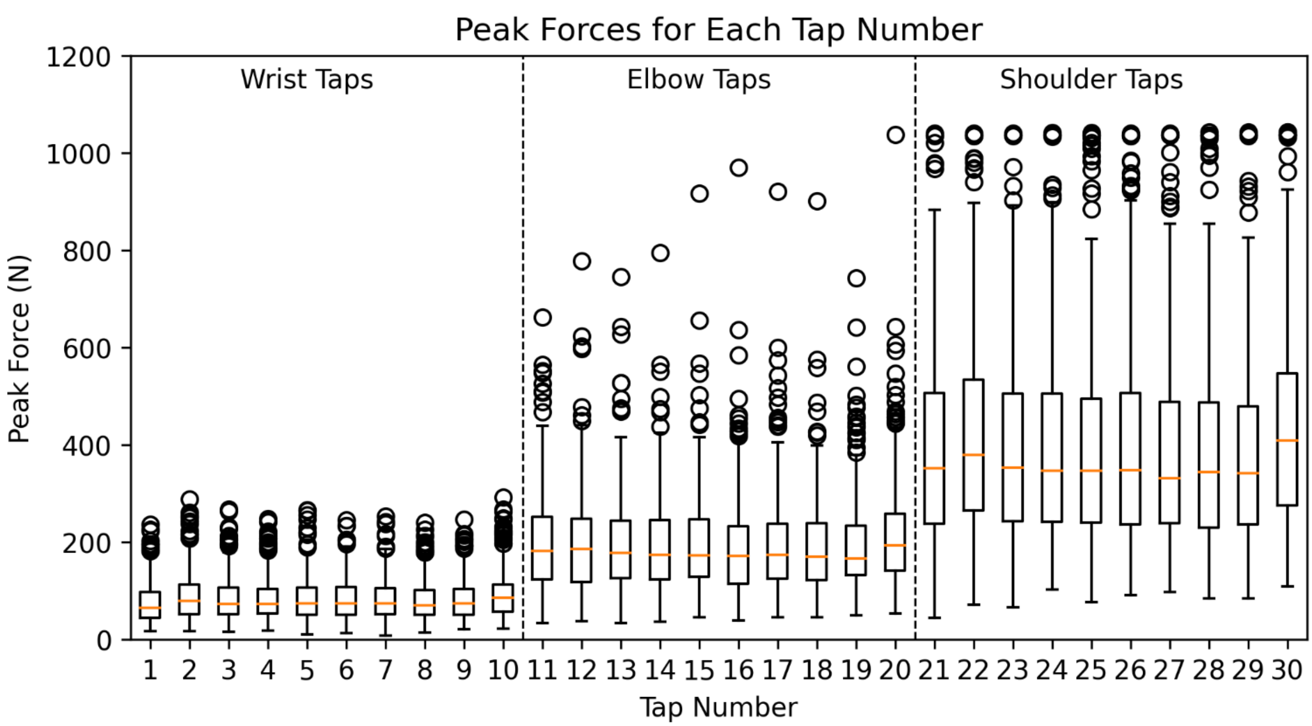


Figure 4: A visualization of the magnitude and variability in peak impact force from the 286 participants from tap 1 to 30. A box plot for each tap number displays the minimum, first quartile, median, third quartile, and maximum values. Outliers are shown using circular symbols. The load cell reaches saturation at 1,000 N, a threshold which was reached in one elbow tap and 75 shoulder taps.

~~4.~~ 3.2 ~~3.2~~ Survey results

There are two important findings from the regression models. ~~.~~First, the information contained in the explanatory variables cannot explain the bulk of the variance in the "tap-o-meter" data. Second, weight and height are significantly and positively correlated with tap force as individual explanatory variables (p-value <= 0.05, Appendix-4~~3~~; p-value <= 0.05, Appendix-5~~4~~); however, the significance is no longer apparent when we include gender (Appendix-6~~5~~; Appendix-7~~6~~). Thus, gender is the only explanatory variable that is significantly correlated with tap force across all multivariate regression models (p-value <= 0.05, Appendix-6~~5~~; p-value <= 0.05, Appendix-7~~6~~).~~.~~

3.3 ~~Due to the poor fit of the models and the small sample size, we only present results from simple bivariate tests (Student t-tests) here. For wrist taps, men on average tap at 88 N while women's average is at 71 N (t = 2.83, p < 0.001). For elbow taps, men on average tap at 205 N while women's average is at 160 N (t = 3.53, p < 0.001). For shoulder taps, men on average tap at 420 N while women's average is at 316 N (t = 4.12, p < 0.01).~~

~~5.~~ ~~3.3~~ Idealization of taps as Gaussian functions

Both the peak force, F_{peak}~~,~~ and loading rate, r~~,~~ are used to idealize the impact curves. First, consider the equation describing a Gaussian function of force, F, as a function of time, t.

$$F(t) = F_{peak} e^{-\frac{1}{2}\left(\frac{t - t_{peak}}{\sigma \cancel{s}}\right)^2}, \quad (3)$$

Where F_{peak} is the peak force and t_{peak} is the time at which the peak force occurs. The duration of the force curve is governed by $\sigma\cancel{s}$, the standard deviation if the Gaussian function were to be describing a normal distribution. Since 99.7% of the curve's magnitude occurs during $6\sigma\cancel{s}$, the duration of impact is defined $6\sigma\cancel{s}$ in our study. Thus, the rise to peak force occurs over approximately $3\sigma\cancel{s}$, leading to the following relationship to calculate the loading rate, r.~~.~~

$$r \cancel{loading\ rate} \approx \frac{F_{peak}}{3\sigma\cancel{s}}, \quad (4)$$

This is an approximation rather than equality because it assumes a linear rise, rather than the non-linear Gaussian shape. However, since loading rate and peak force are the two metrics ascertained from the measured data, this approximation provides a convenient way to idealize the measured force curves. Rearranging the approximation yields

$$\sigma \cancel{s} \approx \frac{F_{peak}}{3 \cancel{*} r \cancel{(loading\ rate)}}\,, \qquad (5)$$

And substituting this relationship for $\sigma \cancel{s}$ in Eq. (3) yields the Gaussian approximation used to idealize the measured force-time curves.

$$F(t) \approx F_{peak} e^{-\frac{1}{2}\left(\frac{3 r \cancel{*} (t - t_{peak}) \cancel{(loading\ rate)}}{F_{peak}}\right)^2}\,, \qquad (6)$$

Using the median metrics along with their 25th and 75th percentiles are plotted in Figure 5.

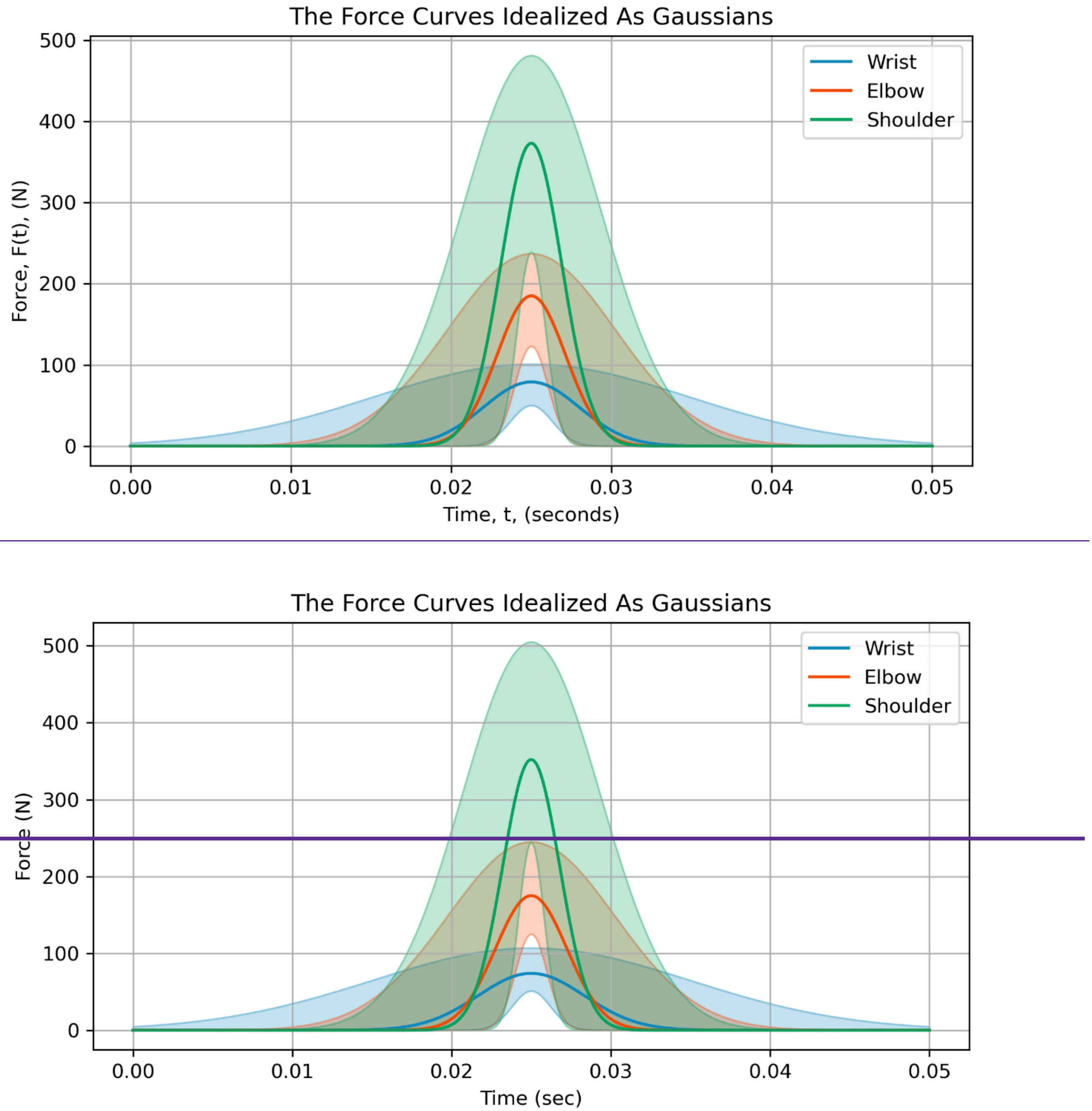


Figure 5: An idealization of the taps as Gaussian functions. The center lines are from the median metrics and the shading is generated from the 25th and 75th percentiles.

By idealizing these tap curves as Gaussians, their respective linear impulsess (i.e. momentums) can be compared by calculating the area under curve (Hibbeler, 2010). Using NumPy's implementation of the trapezoidal rule (Harris et al., 2020), the median wrist, elbow, and shoulder tap impulses are 0.62, 0.99, and 1.58 N*s, respectively.

4. Discussion

4.1 Methods

4.1.1 The tap-o-meter

The tap-o-meter was initially developed using parts in stock at the Norwegian Water Resources and Energy Directorate (NVE). Early testing suggested that a 50 kg load cell which NVE had in stock would be capable of accurately recording the impact force from taps. Based on data collected prior to those showcased in this paper, it became evident that the impact forces from some participants plateaued around 600 N (~60 kg) on their shoulder taps. This level surpassed the recommended operating range of the load cell but stayed within the ultimate overload capacity (150 kg). We pinpointed the problem to the amplifier, which was reaching its saturation point.

We considered the amplifier properties to avoid two potential issues. Setting it too high would mean losing detail in measuring light wrist taps due an increased background noise. On the other hand, setting it too low would make it impossible to measure the strongest impact forces.

To address this, we developed a new adjustable amplifier that we tuned to a range from 5 to 1,000 N. This calibration aimed to balance the ability to detect high-impact forces while maintaining a low background noise for measuring the force of lighter taps. The defined range stayed safely below the load cell's ultimate overload threshold of 1,225 N. Despite the new adjustment with the amplifier's upper limit set to 1,000 N, saturation still occurred in rare instances: once during elbow-level taps (representing 0.03% of such taps) and 75 times for shoulder-level taps (2.63% of such taps). All of these taps are identified as outliers (Figure 4).

The tap-o-meter was calibrated by stacking steel plates of known mass on the load cell. The calibrated range was limited to 30 kg (294 N) to prevent the masses from tipping over and maintain even load distribution. In our analyses, we are assuming the load cell continues to respond linearly through its working range up to its operating limit. Thus, the upper end of shoulder tap measurements may not be accurate. Furthermore, we are assuming the load cell responds similarly to dynamic load 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 include a load cell with a higher range (e.g. 200 kg), load cells designed for impacts (e.g. piezo-resistive), and a fixture to ensure centered loading. 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). 79185373

4.1.2 Data collection

Initially, our idea was to have a representative group of participants with different levels of training. However, after the first data collection event, we realized that most novices did not know how to do the test, and it was difficult to get a representative sample from less experienced participants.

Each participant was asked to fill out a survey. In retrospect, an estimate of how many ECTs each participant does in a season would be of interest. Most participants noted that they do it regularly at work, recreation or both, but we do not have an idea of how frequently they conduct ECTs.

Furthermore, systematic notes about the tapping technique would also be of interest. A qualitative remark is that many of the participants do not use their fingertips on wrist taps as in the standards (American Avalanche Association, 2022; Canadian Avalanche Association, 2016). There was also a large variability in impact forces as a result of using the weight of the arm versus a shoulder tap so hard that it hurts the hand. In some cases, participants placed a glove on the shove to soften the blow. We also observed that some participants increased their impact force during the ten taps within each level, but we do not see this in our overall data (Fig. 4).

4.1.3 Metric Selection

To gain a quantitative understanding of the collected data, peak force and loading rate were the chosen metrics. Other quantities such as impact duration, rise time, and stress were considered but not chosen. Impact duration was not used because the measurements frequently contained long, oscillatory tails that are an artifact of the load cell rebounding and vibrating – a phenomenon expected to be less present during an actual field test. Rise time was calculated as an intermediary step to loading rate. However, loading rate was chosen because snow's response has been shown to depend on its rate of deformation (Shapiro et al., 1997). Lastly, our measurements are presented as forces (N) rather than stresses (kPa) because presenting it as a stress would rely on an assumption of cross-sectional area.

4.12. Results

Using the data from the "tap-o-meter", we are able tocan 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.

4.1~~2.1~~ Peak applied~~Impact~~ force

If we compare the results from our study with the ones from Sedon (2021) and Griesser et al. (2023), we find surprisingly large discrepancies when comparing the measured mean values (Table 4). It is unlikely that participants from New Zealand (Sedon, 2021) tap half as hard as Griesser et al. (2023) observed or one-third of what we observe in our sample from Scandinavia, Europe, and North America. Griesser et al. (2023) recognize that they are not able to accurately measure peak force values due to their lower sampling rate but that the relative differences are systematic when comparing the mean values from wrist, elbow, and shoulder with data from our study.

Table 4~~x~~: A comparison of mean peak force values for wrist, elbow, and shoulder from relevant studies.

Reference	**Wrist (mean)**	**Elbow (mean)**	**Shoulder (mean)**	**Sampling rate**	**Samples**
Sedon (2021)[1]	24 N	62 N	136 N	Unknown	69
Griesser~~Silke~~ et al. (2023)	41~~-46~~ N	97~~-105~~ N	185~~-200~~ N	100 Hz	62
This study	79 N	185 N	373 N	50 kHz	286

[1] Sedon (2021) uses the maximum value from each ~~tap level~~loading step to calculate the mean between participants.

We ~~have measured~~ estimate the average loading duration of the impact curve to be around 20 ms for the wrist, 15 ms for the elbow and 10 ms for the shoulder (Figure 5). At a sampling rate of 100 Hz, we would only measure the impact force every 10 ms, making it unlikely to capture the peak force value accurately. The discrepancies in sampling rates make for an invalid comparison~~it impossible to compare~~ peak force values between the studies. However, the relative difference between wrist, elbow, and shoulder is almost identical for all studies. All three studies have an approximately doubling in peak impact force from wrist to elbow to shoulder.

~~When interpreting the descriptive definitions from each tap level, it is impossible to infer which impact forces should be used as a baseline for each tap level. 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, in Canada, there is no description of how hard each tap should be other than that it should not hurt at shoulder level (Canadian Avalanche Association, 2016). However, this would depend on the participant's pain tolerance, snow properties (dampening) and the participant's glove thickness. The iImpact 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"). Furthermore, 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 tap level. 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).~~

4.12.2 Variability between participants

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).

Griesser et al. (2023) highlighted the differences as a positive outcome of the test and that impact forces are somewhat reli able. Even though each loading steps are statistically different, it is not appropriate to use the average results in individual cases, especially in scenarios with the potential for fatal outcomes. The main difference in our argument lies in the inherent risk of relying solely on mean statistics in avalanche terrain, which is a risky environment. The presence of significant overlap between the 25th-75th percentile ranges of force applied during elbow taps with those of wrist and shoulder taps, where ~18% and ~26% of the data for elbow taps overlap with wrist and shoulder taps, respectively (Appendix-3). These overlaps have practical significance in real-world applications. Our interpretation aligns with the principle of ‘err on the side of caution,’ especially in fields where the consequences of errors can be catastrophic.

In our study involving 286 participants, we observe different mean and median values for each tap level. If we consider the interquartile range which represents the data between the 25th and 75th percentile, there is nearly no overlap between tap levels. However, when considering the range of data (excluding outliers), from the minimum to the maximum values (as represented by the whiskers in a Fig. 4), there is a considerable degree of overlap. This suggests that fifty percent of the data is distinctly different for each tap level. On the other hand, that implies that up to fifty percent of the data could overlap with other tap levels. Furthermore, there are examples of participants with harder wrist taps than other participants' shoulder taps. This has implications for the consistency and reliability of the test results.

4.1.3 Body characteristics, gender, and regionSurvey results

Sedon (2021) did not investigate whether there were differences due to weight, height, gender, or geographical region. Griesser et al. (2023) investigated shoulder height and found that participants with greater shoulder height had higher impact forces. They also mention that they found statistically significant correlations when comparing against height and weight, but no p-values are provided. Our main finding from the survey data is that only gender hais statistically significant relationship with peak force. Body features (weight and height) areis also correlated with peak tap force, but when included in a multivariate analysis with gender, they disappear. We believe the correlation found by GriesserSilke et al. (2023) for body features is likely due to men being, in general, taller, and heavier.

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 , potentially resulting from the small sample size in a statistical context (n=286).

4.12.43 Idealization of taps as Gaussian functions

The Gaussian function is often used in wave propagation problems because it represents a smooth, continuous pulse of disturbance (Langtangen & Linge, 2017). The measured shape of force-time curves is not a perfect Gaussian (Fig. 3), particularly after the peak force has been reached. The noisy, oscillatory decay following the peak is attributed, in part, to the instrumentation. Despite these imperfections, we intend to use~~show~~ this idealization as a steppingstone towards mathematical modelling efforts. In addition to providing this steppingstone, the idealization shown in Figure 5 provides a visualization of peak force, loading rate, impact duration, and variability associated with these quantities. The taps from the shoulder are generally a sharper pulse (i.e. shorter duration, higher peak force) than a wrist tap. Despite the impact duration decreasing with increasing load step, there is an increase in linear impulse. The linear impulse is equated to the change in linear momentum of the system (Hibbeler, 2010). Thus, the increase in snow's momentum from a hand tap is expected to be larger for higher load steps despite the shorter duration of impacts. The Gaussian idealization provided a convenient method of comparing linear impulses from the tap data whereas direct numeric integration of the load cell data would be inaccurate due to the long, oscillatory tails.

4.5~~2~~ Future topics of discussion for improved standards

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

4.5~~2~~.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 ~~tapping force and~~ 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 (Canadian Avalanche Association, 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"). ~~A "wisdom of crowds" approach could be taken~~

and the median values in our study could be used, or aAn alternative to the “wisdom of the crowds” is a -selection of experts could choose to define the appropriatese 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 paper 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 (e.g., stuffblock test or known weights), but this option 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 (202443) 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.

4.52.2 Limit the ECT test’s interpretation

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 valid to interpret ECTP20 differently compared to ECTP24 in individual cases, 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 (Simenhois et al., 2018). In this approach, we would consider the test result to be unstable if crack propagation occurs, and stable otherwise. This interpretation raises the question of why having three steps in the loading procedure. If the avalanche community aims to maintain consistency in this three-step loading method, it should adopt a refined version of the standards currently used in the United States, Canada, and Norway.

4.33 Future work

During data collection, we asked participants if they regularly conduct CTs or ECTs for work, recreation or both. Participants were also asked to self-evaluate their avalanche assessment level on a scale from 1 to 6, following the definitions from the CARE-panel study (Hetland & Mannberg, 2023). Our hypothesis was that more experienced participants, particularly those frequently performing stability tests, would be more consistent within each loading step. However, the study''s shift in focus towards more experienced individuals (see Section 4.1.2) meant that we lacked a suitable reference group for comparison. 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.

Snow's response to impact forces remains an active research topic and is out of the scope of this study. However, variability in magnitude and duration of applied force will result in some variability of the stress state within the snow which may lead to variability in test results. For more on this topic, we refer the reader to studies by Napadensky (1964), Wakahama & Sato (1977), Johnson et al. (1993), Schweizer et al. (1995), van Herwijnen & Birkeland (2014), Thumlert & Jamieson (2015), and Griesser et al. (2023), and Verplanck and Adams (2024). Quantifying how variability in the applied force may lead to different ECT results would be a useful extension of our work presented here.

4.64. Limitations

4.64.1 The "tap-o-meter"

While our study has made significant strides in accurately understanding the tap-o-meter's calibration and limitationsobserving the force-time curves from hand taps, there are still areas that require further exploration. For instance,, the upper end of shoulder 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. as we assume the load cell will 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. 2,000 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). (Griesser et al. 2023) and 105 Hz (Thumlert and Jamieson, 2015). Sedon (2021) do not provide any sampling rate for their study.

4.64.2 Data collection

Initially, our idea was to have a representative group of participants with different levels of training. However, after the first data collection event, we realized that most novices did not know how to do the test, and it was difficult to get a representative sample from less experienced participants.

Each participant was asked to fill out a survey. In retrospect, an estimate of how many ECTs each participant does in a season would be of interest. Most participants noted that they do it regularly at work, recreation or both, but we do not have an idea of how frequently they conduct ECTs.

Furthermore, systematic notes about the tapping technique would also be of interest. A qualitative remark is that many of the participants do not use their fingertips on wrist taps as in the standards (American Avalanche Association, 2022; Canadian Avalanche Association, 2016). There was also a large variability in impact forces ~~as a result of~~because of different techniques such as using the weight of the arm versus a shoulder tap so hard that it hurts the hand. In some cases, participants placed a glove on the shove to soften the blow. We also observed that some participants increased their impact force during the ten taps within each level, but we do not see this in our overall data (Fig. 4).

4.7 Future work

During data collection, we asked participants if they regularly conduct CTs or ECTs for work, recreation or both. Participants were also asked to self-evaluate their avalanche assessment level on a scale from 1 to 6, following the definitions from the CARE-panel study (Hetland & Mannberg, 2023). Our hypothesis was that more experienced participants, particularly those frequently performing stability tests, would be more consistent within each loading step. However, the study's shift in focus towards more experienced individuals (see Section 4.1.2) meant that we lacked a suitable reference group for comparison. 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.

Snow's response to impact forces remains an active research topic and is out of the scope of this study. However, variability in magnitude and duration of applied force will result in ~~some~~ variability of the stress state within the snow which may lead to variability in test results. For more on this topic, we refer the reader to studies by Napadensky (1964), Wakahama & Sato (1977), Johnson et al. (1993), Schweizer et al. (1995), van Herwijnen & Birkeland (2014), Thumlert & Jamieson (2015), Griesser et al. (2023), and Verplanck and Adams (2024). Quantifying how variability in the applied force may lead to different ECT results would be a useful extension of our work presented here.

5. ~~6.~~

~~5.~~ Conclusion

In this study, we ~~have~~ developed a device that can accurately measure force-time curves from the hand-tap loading method~~hand taps' peak force and loading rate~~. 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 ~~tap level~~loading step (i.e. wrist, elbow, shoulder). There is nearly no overlap in peak force from the 25^th^ to 75^th^ percentile between ~~tap level~~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.

We investigated whether the differences in weight, height, gender, or geographical region influence peak force. We found that the explanatory variables largely do not account for the variance in the data, and only gender consistently correlates with tap force across all statistical models~~When we investigated whether there were differences due to weight, height, gender, and/or geographical region, the regression models reveal two important findings. First, the information in the explanatory variables cannot explain the bulk of the variance in the tap-o-meter data. Second, gender is the only explanatory variable that is significantly correlated with tap force across all multivariate regression models~~.

Our results provide an answer to the question of "How hard *do* we tap?" but not necessarily "How hard *should* we tap?". We recommend our data be used to facilitate discussions related to updating guidelines for the hand-tap loading method, possibility of including thresholds of peak force and loading rate for each ~~tap level~~loading step, and revisiting the interpretation of test results given the variability of applied load with the current tapping methodology.

~~We have idealized hand taps as Gaussian functions, which could work as a steppingstone towards future mathematical modelling. The idealization also lets us visualize the peak force, loading rate, impact duration, and variability associated with these quantities. Taps from the shoulder are generally sharper pulses (i.e., shorter duration, higher peak force) than wrist taps.~~

~~The main finding from our study is that there is a substantial~~ large ~~variability in impact force among individuals conducting ECTs, even though each tap level is statistically different from each other. The inconsistency challenges the reliability and reproducibility of ECT results, potentially leading to dangerous interpretations in avalanche decision-making, forecasting and risk assessments. 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.Therefore, we recommend updating the standard for the ECT. How the standard should be updated, specifically, should be a decision made collaboratively by the broader avalanche community.~~

5.1 Calls to action

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

5.1.1 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 tapping force and technique. The community will need to agree on what the ideal impact force-time curves are. A “wisdom of crowds” approach could be taken and the median values in our study could be used, or a selection of experts could choose to define these windows. 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 tap level. Another solution could be to develop a tool that ensures consistent impact force (e.g., stuffblock test or known weights).

5.1.2 Limit the test’s interpretation

Another path would be to limit the level of detail in test result interpretation. As members of the international avalanche community, we need to assess whether the number of taps holds significant value in interpreting ECT results. The added complexity from Winkler et al. (2009) and Techel et al. (2020) makes the test more dependent on the individual’s impact force. If we do not deem the number of taps to be important, we could revert to the original, binary interpretation from Simenhois and Birkeland (2006), focusing solely on crack propagation. In this approach, we would consider the test result to be unstable if crack propagation occurs, and stable otherwise. This interpretation merits the question of why having three steps in the loading procedure. If the community aims to maintain consistency in this three-step loading method, it should adopt a refined version of the standards currently used in the United States, Canada, and Norway.

Data availability

The data needed to replicate the study is available in our Open Science Framework repository (Toft et al., 2023).

Author contributions

The study was conceptualized by HT, SV, and ML. HT developed and built the three “tap-o-meters”. All authors actively participated in data collection at various events. SV, with HT’s assistance, conducted the data pre-processing. HT led the analysis on trends and variability among participants, incorporating insights from SV and ML. The conceptualization of taps as Gaussian functions was primarily driven by SV, with inputs from HT and ML. All authors were actively involved in the preparation, editing, and review of the original draft.

Acknowledgements

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 of~~all the study participants as well.

Competing interests

The authors declare that they have no conflict of interest.

References

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

Benedetti, L., Gaume, J., & Fischer, J.-T. (2019). A mechanically-based model of snow slab and weak layer fracture in the Propagation Saw Test. *International Journal of Solids and Structures*, *158*, 1–20. https://doi.org/10.1016/j.ijsolstr.2017.12.033

Birkeland, K. W., & Johnson, R. F. (1999). The stuffblock snow stability test: comparability with the ~~rutschblock~~29utschblock, usefulness in different snow climates, and repeatability between observers. *Cold Regions Science and Technology*, *30*(1–3), 115–123. https://doi.org/10.1016/S0165-232X(99)00015-4

Birkeland, K. W., & Simenhois, R. (2008). The Extended Column Test: Test Effectiveness, Spatial Variability, and Comparison with the Propagation Saw Test. *International Snow Science Workshop, Whistler, British Colombia*, 867–874.

Canadian Avalanche Association. (2016). Observation guidelines and recording standards for weather, snowpack and avalanches. In C. Campbell, D. McClung, B. Jamieson, B. Sayer, R. Whelan, J. Floyer, & S. Garvin (Eds.), *Canadian Avalanche Association* (6th Edition). Canadian Avalanche Association.

Clarkson, P. (1993). Compression test. *Avalanche News 40*, 9–9.

Föhn, P. (1987). The rutschblock as a practical tool for slope stability evaluation. *IAHS Publ*, *162*, 223–228.

Gauthier, D., & Jamieson, B. (2008). Fracture propagation propensity in relation to snow slab avalanche release: Validating the Propagation Saw Test. *Geophysical Research Letters*, *35*(13), 2–5. https://doi.org/10.1029/2008GL034245

Gauthier, D., & Jamieson, J. B. (2006). Evaluating a prototype field test for weak layer fracture and failure propagation. *International Snow Science Workshop, Telluride, Colorado*, 107–116.

Griesser, S., Pielmeier, C., Boutera Toft, H., & Reiweger, I. (2023). Stress measurements in the weak layer during snow stability tests. *Annals of Glaciology*, 1–7. https://doi.org/10.1017/aog.2023.49

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Heierli, J., Gumbsch, P., & Zaiser, M. (2008). Anticrack nucleation as triggering mechanism for snow slab avalanches. *Science*, *321*(5886), 240–243. https://doi.org/10.1126/science.1153948

Hendrikx, J., & Birkeland, K. (2008). Slope Scale Spatial Variability Across Time and Space: Comparison of Results from Two Different Snow Climates. *International Snow Science Workshop, Whistler, British Colombia*, 155–162.

Hetland, A., & Mannberg, A. (2023). *CARE Panel*. UiT The Arctic University of Norway. https://uit.no/research/carepanel

Hibbeler, R. C. (2010). *Dynamics* (12th ed.). Pearson Prentice Hall.

Jamieson, B., & Johnston, C. (1996). The Compression Test for Snow Stability. *International Snow Science Workshop, Banff, Alberta*, 118–125.

Johnson, J. B., Solie, D. J., Brown, Joseph. A., & Gaffney, E. S. (1993). Shock response of snow. *Journal of Applied Physics*, *73*(10), 4852–4861. https://doi.org/10.1063/1.353801

Kellermann, W. (1990). Erfahrungen mit der Norwegermethode und deren Vergleich mit dem Rutschblock/Keil. In W. Kellermann (Ed.), *Vortrag beim Zentralen Kaderkurs Lawinen des SAC*. Swiss Alpine Club.

LaChapelle, E. R. (1980). The Fundamental Processes in Conventional Alavalanche Forecasting. *Journal of Glaciology*, *26*(94), 75–84. https://doi.org/10.3189/S0022143000010601

Laerdal. (2023). *CPRmeter 2 User Guide*. Laerdal. https://cdn.laerdal.com/downloads/f6537/cprmeter_2_user_guide_en

Langtangen, H. P., & Linge, S. (2017). *Finite Difference Computing with PDEs* (Vol. 16). Springer International Publishing. https://doi.org/10.1007/978-3-319-55456-3

McClung, D. (1979). Shear fracture precipitated by strain softening as a mechanism of dry slab avalanche release. *Journal of Geophysical Research: Solid Earth*, *84*(B7), 3519–3526.

McClung, D. M., & Borstad, C. P. (2012). Deformation and energy of dry snow slabs prior to fracture propagation. *Journal of Glaciology*, *58*(209), 553–564. https://doi.org/10.3189/2012JoG11J009

McClung, D., & Schaerer, P. (2006). The Avalanche Handbook. *The Mountaineers Books*, 1–342.

Moner, I., Gavaldà, J., Bacardit, M., Garcia, C., & Martí, G. (2008). Application of Field Stability Evaluation Methods to the Snow Conditions of the Eastern Pyrenees. *International Snow Science Proceedings, Whistler, British Colombia*, 386–392.

Müller, K., Techel, F., Mitterer, C., Feistl, T., Sofia, S., Roux, N., Palmgren, P., Bellido, G. M., & Bertranda, L. (2023). The EAWS Matrix, A Look-Up Table for Regional Avalanche Danger Level Assessment, and its Underlying Concept. *International Snow Science Workshop, Bend, Oregon*, 540–546.

Napadensky, H. (1964). Dynamic response of snow to high rates of loading. *US Army Material Command, Cold Regions Research & Engineering Laboratory*, 1–52.

Norwegian Water Resources and Energy Directorate. (2022). *Felthåndbok for Snø og Skredobservasjoner* (J. Aasen, Ed.; 3rd Edition). Norwegian Water Resources and Energy Directorate.

Perla, R. I., & LaChapelle, E. R. (1970). A theory of snow slab failure. *Journal of Geophysical Research*, *75*(36), 7619–7627. https://doi.org/10.1029/JC075i036p07619

Reiweger, I., Gaume, J., & Schweizer, J. (2015). A new mixed-mode failure criterion for weak snowpack layers. *Geophysical Research Letters*, *42*(5), 1427–1432. https://doi.org/10.1002/2014GL062780

Reuter, B., & Schweizer, J. (2018). Describing Snow Instability by Failure Initiation, Crack Propagation, and Slab Tensile Support. *Geophysical Research Letters*, *45*(14), 7019–7027. https://doi.org/10.1029/2018GL078069

Schweizer, J., & Jamieson, J. (2010). Snowpack tests for assessing snow-slope instability. *Annals of Glaciology*, *51*(54), 187–194. https://doi.org/10.3189/172756410791386652

Schweizer, J., Schneebeli, M., Fierz, C., & Föhn, P. M. B. (1995). Snow mechanics and avalanche formation: field experiments on the dynamic response of the snow cover. *Surveys in Geophysics*, *16*(5–6), 621–633. https://doi.org/10.1007/BF00665743

Sedon, M. (2021). Evaluating forces for extended column tests and compression tests. Canadian Avalanche Journal, 127, 39-41.

Shapiro, L., Johnson, J., Sturm, M., & Blaisdell, G. (1997). *Snow mechanics: review of the state of knowledge and applications*.

Sigrist, C., & Schweizer, J. (2007). Critical energy release rates of weak snowpack layers determined in field experiments. Geophysical Research Letters, 34, https://doi.org/10.1029/2006GL028576

Simenhois, R., & Birkeland, K. (2006). The Extended Column Test: A Field Test for Fracture Initiation and Propagation. *International Snow Science Workshop, Telluride, Colorado*, 79–85.

Simenhois, R., & Birkeland, K. W. (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. https://doi.org/10.1016/j.coldregions.2009.04.001

Surowiecki, J. (2005). *The wisdom of crowds*. Anchor.

Techel, F., Winkler, K., Walcher, M., van Herwijnen, A., & Schweizer, J. (2020). On snow stability interpretation of extended column test results. *Natural Hazards and Earth System Sciences*, *20*(7), 1941–1953. https://doi.org/10.5194/nhess-20-1941-2020

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

Toft, H. B., Verplanck, S. V., & Landrø, M. (2023). *Tap-o-meter data*. Open Science Framework. https://doi.org/10.17605/OSF.IO/BV5PM

Tukey, J. (1977). *Exploratory data analysis* (Vol. 2).

Van Herwijnen, A., & Birkeland, K. W. (2014). Measurements of snow slab displacement in Extended Column Tests and comparison with Propagation Saw Tests. *Cold Regions Science and Technology*, *97*, 97–103. https://doi.org/10.1016/j.coldregions.2013.07.002

van Herwijnen, A., Heck, M., & Schweizer, J. (2016). Forecasting snow avalanches using avalanche activity data obtained through seismic monitoring. *Cold Regions Science and Technology*, *132*, 68–80. https://doi.org/10.1016/j.coldregions.2016.09.014

Verplanck, S., & Adams, Edward. (2024). Dynamic models for impact-initiated stress waves through snow columns. *Journal of Glaciology*. https://doi.org/doi:10.1017/jog.2024.26

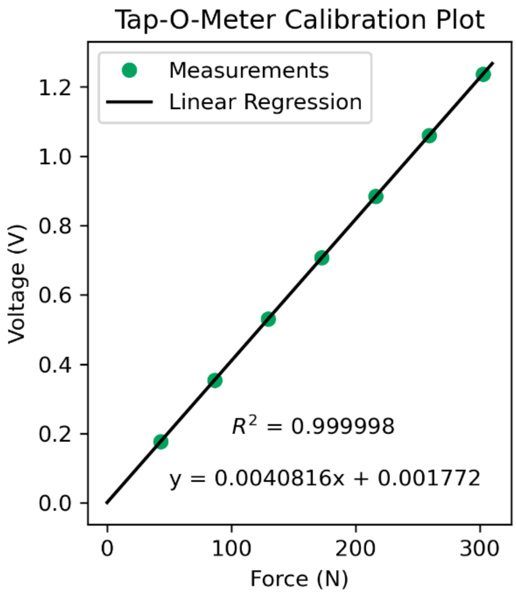
Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … Vázquez-Baeza, Y. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

Wakahama, G., & Sato, A. (1977). Propagation of a Plastic Wave in Snow. *Journal of Glaciology*, *19*(81), 175–183. https://doi.org/10.3189/S0022143000029269

Weißgraeber, P., & Rosendahl, P. L. (2023). A closed-form model for layered snow slabs. *The Cryosphere*, *17*(4), 1475–1496. https://doi.org/10.5194/tc-17-1475-2023

Winkler, K., & Schweizer, J. (2009). Comparison of snow stability tests: Extended column test, ~~rutschblock~~32utschblock test and compression test. *Cold Regions Science and Technology*, *59*(2–3), 217–226. https://doi.org/10.1016/j.coldregions.2009.05.003

Appendix



Appendix-1: The “tap-o-meter”~~tap-o-meter~~ was calibrated using known weights ranging from ~50 to 300 N~~kg~~.

Event	Date	Samples
European Avalanche Warning Services General Assembly	15.06.2022	62
Montana State University Snow~~ki~~ and Avalanche Workshop	26.10.2022	25
Norwegian Avalanche Observer Workshop	08.11.2022	46
UIAGM General Assembly Norway	12.11.2022	27
Friends of the Gallatin National Forest Avalanche Center Instructor Training	15.11.2022	9
Southwest Montana Ski Patrol Snow Science Day	18.11.2022	30
Mountain Guides Meeting, Innsbruck #1	30.11.2022	17
Mountain Guides Meeting, Innsbruck #2	15.12.2022	15
Forecasters at Parks Canada	24.02.2023	4
Colorado Avalanche Information Center	02.03.2023	5
Sawtooth Avalanche Center	08.03.2023	26
Chugach National Forest Avalanche Information Center	13.03.2023	3
Gallatin National Forest Avalanche Center Professional Development Workshop	05.04.2023	17

Appendix-2: A description of each event, date and number of samples gathered.

	< Wrist (< 50 N)	Wrist (50-112 N)	Elbow (112-238 N)	Shoulder (238-481 N)	> Shoulder (> 481 N)
Wrist	23.48%	53.79%	21.71%	1.02%	0.00%
Elbow	0.92%	17.79%	53.82%	25.75%	1.73%
Shoulder	0.04%	1.30%	22.24%	48.70%	27.72%

Appendix-3: To showcase the overlap between loading steps, we have made a confusion matrix based on a tapping norm. The IQR for wrist, elbow and shoulder is respectively 50-101 N, 123-237 N and 239-481 N. We have selected the value between the highest IQR value and lowest IQR value in each loading step to define the tapping norm. Using these values, we can make a confusion matrix to highlight how many hand taps that are within each interval. From this, we can e.g. see that 17.79% of elbow taps are within the wrist tapping norm, or 25.75% is within the shoulder norm.

Model 1: Weight

	ln(wrist)	ln(elbow)	ln(shoulder)
Weight	0.008**	0.006**	0.006*
	(0.002)	(0.002)	(0.002)
Region (reference is European Alps)			
North America	0.085	0.040	0.095
	(0.068)	(0.059)	(0.064)
Scandinavia	-0.041	-0.174*	-0.084
	(0.080)	(0.068)	(0.070)
Constant	3.718**	4.717**	5.426**
	(0.187)	(0.179)	(0.183)
N	286.000	286.000	286.000
F-value	4.809	5.294	3.649
R2-adjusted	**0.032**	**0.051**	**0.033**
AIC	424.658	359.685	388.486

Appendix-42: OLS for weight. P-values: + p <= 0.1, * p <= 0.05 and ** p <= 0.01.

Model 2: Height

	ln(wrist)	ln(elbow)	ln(shoulder)
Height	0.011**	0.010**	0.008*
	(0.004)	(0.003)	(0.003)
Region (reference is European Alps)			
North America	0.106	0.058	0.111+
	(0.068)	(0.059)	(0.064)
Scandinavia	-0.031	-0.167*	-0.076
	(0.079)	(0.067)	(0.068)
Constant	2.335**	3.444**	4.477**
	(0.637)	(0.588)	(0.587)
N	286.000	286.000	286.000
F-value	4.292	6.406	3.800

R2-adjusted	**0.035**	**0.057**	**0.032**
AIC	423.625	357.627	388.669

Appendix-53: OLS for height. P-values: + p <= 0.1, * p <= 0.05 and ** p <= 0.01.

Model 31: Gender and weight

	ln(wrist)	ln(elbow)	ln(shoulder)
Female	-0.137	-0.218**	-0.256**
	(0.089)	(0.080)	(0.089)
Weight	0.005+	0.002	0.001
	(0.003)	(0.003)	(0.003)
Region (reference is European Alps)			
North America	0.100	0.065	0.123+
	(0.068)	(0.057)	(0.063)
Scandinavia	-0.023	-0.146*	-0.051

	(0.080)	(0.068)	(0.071)
Constant	3.933**	5.061**	5.829**
	(0.222)	(0.209)	(0.255)
N	286.000	286.000	286.000
F-value	4.007	5.835	6.517
R2-adjusted	**0.036**	**0.070**	**0.058**
AIC	424.503	354.705	381.780

Appendix-642: OLS for gender and weight. P-values: + p <= 0.1, * p <= 0.05 and ** p <= 0.01.

Model 42: Gender and height

	ln(wrist)	ln(elbow)	ln(shoulder)
Female	-0.121	-0.199*	-0.266**
	(0.095)	(0.088)	(0.091)
Height	0.008+	0.004	0.000
	(0.004)	(0.004)	(0.004)
Region (reference is European Alps)			

North America	0.113+	0.070	0.126*
	(0.067)	(0.057)	(0.063)
Scandinavia	-0.018	-0.146*	-0.048
	(0.079)	(0.066)	(0.069)
Constant	2.955**	4.470**	5.847**
	(0.801)	(0.742)	(0.801)
N	286.000	286.000	286.000
F-value	3.782	6.177	6.519
R2-adjusted	**0.037**	**0.072**	**0.058**
AIC	424.046	354.146	381.846

Appendix-75~~3~~: OLS for gender and height. P-values: + p <= 0.1, * p <= 0.05 and ** p <= 0.01.