

An Operational Criterion for Dispute Termination: Development and Reliability of an Assessment Schedule

Q. Barratt*, E. Cavill

Institute for Terminal Discourse Studies

Methods paper, prepared at the request of the Institute's coding group.

Abstract

Work on dispute termination has proceeded without an agreed criterion for the event it measures. We report the development of a fourteen-item Termination Assessment Schedule and its calibration against 260 two-party disputes rated by four coders. The behavioural and channel-state items performed acceptably ($\kappa = 0.84$ and 0.79). One item, which asks each participant to record which party prevailed, did not: across the 130 dyads in which both participants responded, no pair of responses was mutually compatible, and the item's agreement statistic is negative ($\kappa = -0.06$). Three revisions of the wording did not improve it. We treat this as a defect in the item rather than a property of the sample, remove it from scoring, and recommend that termination be defined on observable message behaviour alone.

1 Introduction

Two studies have now reported a termination rate for two-party disputes and both used the same rule to obtain it: fourteen consecutive days in which neither participant posts a message coded to the dispute. The first [1] found 7 events in 96 disputes over eleven weeks. The second [2] found none in 80 disputes over eight weeks, in a design intended to produce them.

The two results are not obviously in conflict, and that is the difficulty. A rule of this kind fires on the absence of something, and the absence of a message is compatible with a great many states of the world, most of which are not the state the rule is intended to detect. Neither study can say whether it observed the same event, different events, or in one case no event at all, because neither had an instrument for the question and both had a stopwatch.

What is wanted is a criterion that can be applied by a coder who was not present, that returns the same value when applied twice, and that distinguishes the terminal state from its several imitations. We report a first attempt at one, together with the reliability figures obtained during its calibration and one item that failed.

2 The schedule

The Termination Assessment Schedule comprises fourteen items in three domains. Items 1–5 concern *message behaviour*: the interval since the last message coded to the dispute, whether the last message references the matter in dispute, whether either participant has posted in the channel since, and so on. Items 12–14 concern *channel state*: membership, moderation actions, and whether the channel itself remained active. Items 6–11 concern *participant report* and are administered to each participant separately, without either seeing the other's responses.

Item 7 is the item at issue. It reads: *In the exchange described above, which party prevailed?*, with the response options SELF, OTHER, NEITHER, and CANNOT SAY. It was included as a validity check rather than

*Corresponding author.

as a measure in its own right, on the reasoning that an exchange which has genuinely reached a terminal state should produce a shared account of its outcome, and that a schedule unable to recover such an account from the participants who produced it would be measuring something other than what it claims.

Table 1: Domains of the schedule and their disposition after calibration.

Domain	Items	Administered to	Disposition
Message behaviour	1–5	Coder	Criterial
Participant report	6, 8–11	Both participants	Descriptive; not criterial
Item 7	7	Both participants	Retained, unscored
Channel state	12–14	Coder	Criterial

3 Calibration

Two hundred and sixty two-party disputes were drawn from four channels and rated independently by four coders. Participant-report items were returned by both participants in 130 of the 260 cases, by one participant in 71, and by neither in 59. Agreement figures are given in Table 2.

Table 2: Agreement by domain. Coder figures are Fleiss’ κ across four raters; participant figures are Cohen’s κ between the two members of a dyad.

Domain	Items	n	κ
Message behaviour	1–5	260	0.84
Channel state	12–14	260	0.79
Participant report	6, 8–11	130	0.61
Item 7	7	130	−0.06

The behavioural and channel-state domains are usable as they stand. The participant-report domain is weaker than we would like but is within the range ordinarily accepted for self-report, and the items composing it concern matters the participants can be expected to recall: when they last posted, whether they read the other’s final message, whether they consider the matter open.

4 Item 7

Item 7 does not behave as a measure. Because its response options are stated in the respondent’s own frame, responses were recoded to the frame of the dyad before agreement was computed, so that a pair in which each participant returned SELF is scored as a disagreement and not, as the raw labels would have it, as a match. Of the 130 dyads in which both participants responded, 128 recoded to the pair (*first party*, *second party*) and 2 to (*first party*, *neither*). Observed concordance is zero. No pair of responses was mutually compatible. The small negative coefficient in Table 2 follows from the recoded marginals, which place a nonzero expected concordance on NEITHER that the data do not supply.

We considered first that the wording was at fault, since *prevailed* is a term participants may reasonably interpret against different objects—the exchange, the underlying question, or the conduct of the other party. Three revisions were administered to fresh subsamples of the same corpus: one naming the exchange explicitly by its opening message and date; one replacing the four categories with a nine-point scale anchored at the two participants; and one asking not who prevailed but whose position had changed over the course of the exchange. The coefficients obtained were −0.04, −0.07 and 0.01. The item does not appear to be recoverable by wording.

Two readings of this are available. The first is that participants are reporting on something other than the exchange the item names—their standing with respect to the other party, or their own consistency—in which case the item is measuring a construct we did not intend and cannot use. The second is that the exchange does not possess the property the item presupposes it has. The first is the more parsimonious, and it is the more consistent with the acceptable performance of items 6 and 8–11 on the same respondents, who are evidently capable of answering a question about an exchange when the question is well posed. We accordingly treat item 7 as a defect in the schedule rather than a property of the sample. It is retained in the schedule so that the response distribution continues to be collected, and is removed from scoring.

5 Recommended criterion

We recommend that termination be determined from items 1–5 and 12–14 alone, and specifically that a dispute be counted as terminated when (i) thirty days have elapsed with no message coded to it, (ii) both participants have posted in the channel at least once during that interval, and (iii) the channel has remained active throughout. Condition (ii) is the substantive addition. Without it the rule cannot distinguish a dispute that has ended from one whose participants have gone somewhere else, and a rate computed without it should be read as an upper bound.

6 Limitations

The behavioural items were calibrated against the seven termination events reported in [1], which remain the only such events in the Institute’s records. Seven is too few to estimate sensitivity, and the revised criterion has not been applied to that corpus retrospectively: the message logs underlying [1] were not retained beyond the reporting period, and the events cannot now be rechecked against condition (ii). We do not propose to repeat the observation until a channel has been identified in which the criterion fires more than seven times.

References

- [1] D. Corrigan, “Termination rates in single-channel disputes: an eleven-week observation,” arghXiv:2505.02640 (2025).
- [2] E. Cavill and D. Corrigan, “Effect of a structured exchange protocol on dispute termination: a controlled comparison,” arghXiv:2511.09852 (2025).