

eConsult Specialist Quality of Response (eSQUARE): A novel tool to measure specialist correspondence via electronic consultation

Journal of Telemedicine and Telecare
2022, Vol. 28(4) 280–290
© The Author(s) 2021



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1357633X21998216
journals.sagepub.com/home/jtt



Christopher Tran^{1,2} , Douglas Archibald^{3,4},
Susan Humphrey-Murto^{1,5}, Timothy J Wood⁵, Nancy Dudek¹,
Clare Liddy^{3,4} and Erin Keely^{1,2}

Abstract

High-quality correspondence between healthcare providers is critical for effective patient care. We developed an assessment tool to measure the quality of specialist correspondence to primary care providers (PCPs) via electronic consultation (eConsult), where specialists provide advice without specialist-patient interactions. We incorporated fourteen previously described features of high-quality eConsult correspondence into an assessment tool named the eConsult Specialist Quality of Response (eSQUARE). Six PCPs and two specialists applied the 10-item eSQUARE tool to 30 eConsults of varying quality as informed by PCP survey data. Content, response process, and internal structure validity evidence was gathered. Psychometric properties were calculated using descriptive statistics and generalizability analyses. Mean total score for low-quality eConsults ($M = 24 \pm 5.6$) was significantly lower than moderate-quality eConsults ($M = 38 \pm 4.7$; $p < 0.001$) which was significantly lower than high-quality eConsults ($M = 46 \pm 3.0$; $p = 0.002$). Reliability measures were high, including generalizability coefficient (0.96), inter-item (≥ 0.55) and item-total correlations (≥ 0.68). A decision study demonstrated that a single rater was adequate to achieve a reliability measure of ≥ 0.70 . This study demonstrates initial validity evidence including multiple reliability measures for the eSQUARE. A single rater is adequate to achieve reliability measures for formative feedback. Future studies can apply the eSQUARE when planning educational initiatives aiming to improve specialist-to-PCP correspondence via eConsult.

Keywords

Electronic consultation, documentation, assessment, medical education, telehealth

Date received: 3 February 2021; Date accepted: 6 February 2021

Introduction

Effective communication between specialists and primary care providers (PCPs) is essential for coordinated patient care. While improved communication may lead to favourable patient outcomes, poor inter-provider correspondence may cause delays in patient care and compromise relationships between clinicians.^{1–3}

An increasingly implemented method for communication among clinicians is electronic consultation (eConsult) – asynchronous communication through a secure electronic platform where PCPs can receive

¹Department of Medicine, University of Ottawa, Canada

²Ottawa Hospital Research Institute, The Ottawa Hospital, Canada

³Department of Family Medicine, University of Ottawa, Canada

⁴Bruyère Research Institute, Canada

⁵Department of Innovation in Medical Education, University of Ottawa, Canada

Corresponding author:

Christopher Tran, Division of Endocrinology and Metabolism
Department of Medicine, University of Ottawa 1967 Riverside Drive,
Room 4-74A Ottawa, ON, K1H 7W9, Canada.

Email: chtran@toh.ca

specialist advice without their patients meeting face-to-face with the specialist.⁴⁻⁷ With more than 50 distinct services worldwide, eConsult is officially endorsed as a standard of practice by both primary care and specialist national regulatory bodies to improve PCP access to specialist advice.^{8,9}

As eConsult becomes more widespread, it is imperative to ensure effective communication when using eConsult. Although both PCPs and specialists report better inter-provider communication with eConsult when compared to conventional letter correspondence,¹⁰⁻¹² PCPs can also have negative experiences with specialist communication via eConsult, particularly if advice is neither clear nor actionable.^{13,14} One way to improve the PCP experience would be to have users rate the quality of specialist correspondence using a formal assessment tool. Feedback generated from this tool could inform faculty development initiatives that aim to improve specialist-to-PCP communication via eConsult.

Existing tools assessing inter-provider correspondence focus primarily on traditional reply letters following face-to-face encounters and do not capture elements unique to eConsult.¹⁵⁻¹⁹ In eConsult, there is no direct patient–specialist interaction, and so the specialist relies entirely on the PCP to provide relevant details regarding history, physical examination and investigations. Also, the PCP is solely responsible for determining if eConsult advice should be implemented, further emphasising the importance of ensuring high-quality specialist communication so that advice can be followed as intended.²⁰

This study is the second of a two-phase project aiming to develop and demonstrate initial validity evidence for a formal assessment tool to measure the quality of specialist communication via eConsult. The first phase used the nominal group technique (consensus group methodology) where an expert panel of 11 high-end eConsult users generated an initial list of 50 items which were then refined to 14 key elements of high-quality eConsult correspondence.²⁰

Methods

Setting

We conducted our study from 2019 to 2020 using eConsults submitted through our regional service (Champlain BASETM, Ottawa, Canada). Established in 2010, our eConsult service has completed more than 65,000 eConsults across 135 specialty groups.

Tool development

Our team of five physicians (four specialists and one PCP) and two health professions education researchers held a meeting to review and refine the 14 key elements for effective eConsult specialist communication, as reported previously.²⁰ Items where authors felt there was overlap in content were combined. The authors also discussed how best to format the assessment tool, for example the number of points on each scale and their descriptive anchors.

To gather sources of validity evidence for our assessment tool, we applied modern validity theory as a framework where five types of validity are postulated: content, internal structure, response process, relation to other variables and consequences.²¹ Our study addresses the first three features using a similar approach as other assessment tools.²²⁻²⁴

Defining eConsult quality

We sought a range in the quality of the eConsults to see if raters could discriminate low- from high-quality eConsults when applying the assessment tool. eConsults submitted between October 2016 and December 2017 were stratified into three quality groups: low, medium and high. Quality was inferred from PCP responses to mandatory close-out surveys where they rate on a five-point scale how helpful and/or educational the specialist response was in guiding ongoing patient management. PCPs also have the option to provide free-text comments. We classified an eConsult as low quality if the PCP assigned a helpfulness rating of 1 or 2 while providing specific comments describing why the eConsult was not helpful. Medium quality eConsults were ones with helpfulness ratings of 3 or 4 and absent or non-specific free-text comments. We defined a high-quality eConsult as one where the PCP assigned the maximum score of 5 for helpfulness and provided positive, detailed comments explaining why the eConsult was particularly helpful. Example free-text comments depicting each quality category are shown in Table 1.

Sample size calculations

Although power analyses are conventional in quantitative research, one cannot be done for assessment tools before they are fully developed.²⁵ Instead, sample size calculations during preliminary testing can be guided by the anticipated feasibility of using the tool and practical considerations such as finances and time.^{26,27} We thus opted for 10 eConsults for each quality category (30 in total), and we aimed to recruit 10 raters to apply the assessment tool to each eConsult.

Table 1. Selection of low-, medium- and high-quality eConsults for eSQUARE testing based on PCP survey data and free-text comments.

eConsult quality	Survey Q3 ^a	Survey Q5 ^b	Example Q5 responses
Low	1 or 2	Negative comments specifically delineating why the eConsult was not helpful	<p>'I would have appreciated a rationale for a particular course of action. Although I can follow directions, I won't be better informed for next time'.</p> <p>'Would appreciate more information about the length of course of both medications, how to administer topical medication, follow up and other options if not initial recommendations working'.</p> <p>'There was no relevant advice given in this consultation. No comments were made relating possible diagnosis, testing or treatment'.</p>
Medium	3 or 4	Absent or non-specific comments	<p>'Again, very helpful!'</p> <p>'Thank you for this timely response. I will be implementing your advice'.</p> <p>'Many thanks for your helpful advice'.</p>
High	5	Positive comments detailing why the eConsult was helpful	<p>'This is an excellent response, clear and with steps to follow and when to refer clearly indicated'.</p> <p>'Fantastic response. Very thorough with excellent, detailed next steps. Really appreciate the time you took to prepare the response. It's very helpful, and I can apply this to other patients as well'.</p> <p>'This was an excellent response, very helpful – gave me a specific plan, and when to refer to the specialist. Super helpful – thanks!'</p>

^aHow helpful and/or educational was this response in guiding your ongoing evaluation or management of the patient?

^bWe would value any additional feedback you provide. [Comments for the specialist will be forwarded to her/him.]

eSQUARE: eConsult Specialist Quality of Response; PCP: primary care provider.

eConsult selection for testing

Among our cohort of 3324 eConsults, only 2% were assigned helpfulness ratings of 1 or 2; some specialties did not have any low-quality eConsults. No more than two eConsults were chosen from a given specialty, except the most popular specialty (dermatology) where three eConsults were included (one for each quality category). One author (C.T.) selected the 30 eConsults and presented them to the research group. We reached consensus that these were representative of the spectrum of quality across the most popular specialties. We removed all personal identifying information, including patient, PCP and specialist names.

Participants

We sent email invites to PCPs considered high-volume eConsult users, that is, those who submitted the above-median number of eight eConsults over a one-year period. Participants in the nominal group phase were excluded. We specifically recruited specialists in endocrinology and obstetrics/gynaecology, since they represent two of our four most frequently requested specialties; dermatology and haematology were the

other top specialties, but they had already participated in the nominal group phase.

Rating process

The 30 de-identified eConsults, each containing the PCP's clinical query and the specialist's response, were provided to participants in a random order. We set up an online rating platform and provided PDF instructions on how to navigate the process, including descriptions of each item on the assessment tool to reduce any ambiguity (Appendix 1). We asked participants to do practice runs with the first few eConsults to familiarise themselves with the assessment tool. Once comfortable with the online platform, they then submitted their ratings. Participants were able to complete ratings in any order and in as many separate sessions as needed.

Statistical analysis

Descriptive statistics, including means for each item along with their standard deviations and ranges, were calculated by first averaging the ratings across raters. Total scores for each eConsult were calculated by

summing the mean ratings received for each of the 10 items on the assessment tool. Any rating assigned as 'not applicable' (N/A) was replaced with the mean of ratings of remaining items assigned by the rater that had assigned the N/A rating. The internal structure of the tool was assessed using both inter-item correlations and item total score reliability; these were calculated using IBM SPSS Statistics for Windows v26 (IBM Corp., Armonk, NY). A generalisability study was felt appropriate for determining reliability measures of our assessment tool, since one can determine how much variance in scores is due to differences between eConsults versus others variables (raters, scale items) along with the interactions with each other.²⁸ Repeated-measures analysis of variance (ANOVA) was conducted with raters, and eSQUARE items were treated as within-subject factors crossed with individual eConsults using G_String V and urGENOVA software platforms.^{29–31} A decision study was conducted to explore how many raters would be needed to produce highly reliable ratings.³² To determine whether the assessment tool could differentiate among low-, medium- and high-quality eConsults, total scores were analysed by conducting a between-subjects ANOVA, with eConsult quality (low, medium and high) treated as a between-subjects factor.

Parametric analyses such as ANOVA are robust and can maintain sufficient validity when applied to Likert scale data, even when some assumptions are violated.^{33,34} We thus felt it was appropriate to apply parametric analyses to the sets of ordinal ratings across eSQUARE items.

Results

Tool development

Items generated from the nominal group session where authors felt there was overlap in content were combined, for example 'Educational: interaction is a learning experience' was combined with 'Separate advice for immediate action and additional material for future reference', and 'Openness to further communication, dialogue' was combined with 'respectful, supportive tone'. 'Timeliness' was removed, since response time is captured electronically by the eConsult platform. The result was a set of 10 items. The authors then reached consensus to use five-point rating scales and a global rating scale to balance rater convenience with assessment tool sensitivity. Scale anchors ranged from 1='not at all' to 5='exemplary'. The assessment tool was named the eConsult Specialty Quality of Response (eSQUARE; Figure 1).

Applying the eSQUARE to eConsults

Participants. Among the 295 PCPs who met the inclusion criteria, 11 agreed to participate; five family physicians and one nurse practitioner completed the study. The two invited specialists completed the study. Thus, a total of eight participants completed eSQUARE ratings for each of the 30 eConsults.

eConsults. When comparing eConsults grouped as low, medium and high quality based on PCP survey data, there was a statistically significant effect of quality ($F(2, 27)=57.7; p<0.001, \eta_p^2=0.81$), suggesting adequate variability among selected eConsults.

Descriptive statistics. Descriptive statistics for eSQUARE items rated on the five-point scale along with item total correlations and frequency N/A ratings are shown in Table 2. While mean ratings for each item were moderately high, raters were also willing to use the full scale, as seen by the range in ratings.

Internal structure measures. Item total score correlations were high, with all but one item having values >0.90 . Inter-item correlations ranged from 0.44 to 0.94, with several items being highly correlated with one another (e.g. items 1–5, item 8 and item 10), indicating potential redundancy of items (Table 3). Since total scores and global rating scores were highly correlated (0.98), subsequent analyses focused only on scale items.

Variance components associated with administering the eSQUARE to eConsults are shown in Table 4. Differences among individual eConsults accounted for most of the variance in the scores (48%), suggesting that ratings varied from one eConsult to another. The next largest variance component other than overall error was the interaction between forms and raters (15%), indicating that raters were somewhat varied in the ratings they assigned to each eConsult. Individual raters accounted for only 1% of the overall variance, suggesting minimal differences across raters. Variance components associated with items tended to be low, that is, items were not a major source of variance. The G-coefficient for the scale with eight raters was 0.96, suggesting that scores from one rater can be generalised to another with high consistency.

To estimate the minimal number of raters needed to optimise score consistency and calculate inter-rater reliability, a decision study was performed (Appendix 2). A single rater applying all 10 eSQUARE items to assess an eConsult response would produce a reliability of 0.74; values >0.7 are considered adequate for formative feedback purposes.³⁵

We are looking to improve communication between health care providers who use eConsult. The purpose of this scale is to provide feedback to specialists on the quality of the advice they provide to health care providers through eConsult.

	Not applicable	Not at all		Acceptable		Exemplary
Scale:	N/A	1	2	3	4	5

General

I feel confident that the specialist advice:

- 1) Is **current (e.g. up-to-date)** and evidence-based where applicable.
- 2) Is **helpful and/or educational**, e.g. the specialist **provides rationale** or evidence for recommendations, or attaches additional material such as guidelines or clinical pathways.
- 3) Is **patient-specific** by incorporating available patient information to the clinical question(s).
- 4) **Addresses each question/concern** posed by the referring provider.

Management

The specialist provides:

- 5) **Specific recommendations** including cost and availability of tests and treatment, while acknowledging if suggestions are off-label. Medication advice includes how to dose, titrate, and monitor response to therapy.
- 6) **Anticipatory guidance** including red-flags (e.g. key features that would prompt further workup) and what to try next if recommendations do not result in a favourable outcome.
- 7) Specific advice as to **when a face-to-face referral would be indicated**.
- 8) **Action items that are doable** with references to local resources where appropriate.

Communication

The specialist advice is:

- 9) **Clear and organized** using logical flow where key information is easy to find.
- 10) Delivered in a **professional, supportive tone** (e.g. specialist acknowledges a difficult and/or complex case) that is open to back-and-forth communication.

Global Rating

- 11) How would you rate the overall quality of the specialist's advice?

Figure 1. The eConsult Specialist Quality of Response (eSQUARE) assessment tool.

Table 2. Descriptive statistics for the eSQUARE assessment tool.

eSQUARE item	Rating ^a		Range ^b		Item total correlation	N/A ratings
	M	SD	Min	Max		
Current	3.9	0.8	1.8	4.9	0.93	1
Educational, provides rationale	3.7	1.1	1.1	5.0	0.96	0
Patient specific	4.0	0.9	1.6	5.0	0.90	1
Addresses each question	3.7	1.1	1.4	5.0	0.91	2
Specific recommendations	3.7	1.1	1.8	5.0	0.95	1
Anticipatory guidance	3.0	1.2	1.0	5.0	0.93	7
When face-to-face referral needed	2.8	1.4	1.0	5.0	0.68	19
Doable action items	3.8	1.1	1.3	5.0	0.97	1
Clear, organised	3.8	1.0	2.0	5.0	0.96	0
Professional, supportive	3.7	1.2	1.0	5.0	0.93	0
Global rating	3.5	1.2	1.0	4.9		

^aEach item was rated on a five-point scale, ranging from 1='not at all' to 5='exemplary'.

^bRange of actual ratings for each item. Non-integer values occurred when N/A ratings were replaced with mean ratings each specific rater assigned to remaining items.

Table 3. Item correlations for the eSQUARE assessment tool.

Item	Description	1	2	3	4	5	6	7	8	9	10
1	Current	1.00	0.98	0.89	0.89	0.90	0.84	0.55	0.93	0.92	0.92
2	Educational, provides rationale	0.98	1.00	0.91	0.92	0.93	0.86	0.61	0.94	0.94	0.94
3	Patient specific	0.89	0.91	1.00	0.83	0.84	0.85	0.64	0.88	0.86	0.87
4	Addresses each question	0.89	0.92	0.83	1.00	0.90	0.84	0.61	0.91	0.91	0.85
5	Specific recommendations	0.90	0.93	0.84	0.90	1.00	0.88	0.63	0.96	0.97	0.91
6	Anticipatory guidance	0.84	0.86	0.85	0.84	0.88	1.00	0.83	0.91	0.88	0.86
7	When face-to-face referral needed	0.55	0.61	0.64	0.61	0.63	0.83	1.00	0.67	0.63	0.61
8	Doable action items	0.93	0.94	0.88	0.91	0.96	0.91	0.67	1.00	0.95	0.92
9	Clear, organised	0.92	0.94	0.86	0.91	0.97	0.88	0.63	0.95	1.00	0.92
10	Professional, supportive	0.92	0.94	0.87	0.85	0.91	0.86	0.61	0.92	0.92	1.00

Table 4. eSQUARE variance components.

Variable	Variance component	% Variance	Description
e	0.96	48%	Variance due to differences among eConsults
r	0.03	1%	Variance due to differences among raters
i	0.12	6%	Variance due to differences among eSQUARE items
er	0.31	15%	Variance due to rater inconsistency across eConsults
ei	0.17	9%	Variance due to eConsult inconsistency across items
ri	0.04	2%	Variance due to rater inconsistency across eSQUARE items
eri	0.38	19%	Overall error and variance due to the interaction between forms, raters and items

e: eConsult; r: raters; i: scale-rated items.

Response process evidence. An aspect of response process evidence is that raters are consistent in how they apply their ratings. Post hoc *t*-tests revealed that the mean \pm standard deviation total score for low-quality eConsults (24 ± 5.6) was significantly lower than moderate-quality eConsults (38 ± 4.7 ; $p < 0.001$) which in turn was significantly lower than high-quality eConsults (46 ± 3.0 ; $p = 0.002$), suggesting that raters using the eSQUARE could reliably distinguish between eConsults divided into three levels of quality.

Discussion

To ensure high-quality communication through emerging health-care innovations such as eConsult, tools to provide guidance, feedback and formal assessment of specialist-to-PCP communication are needed. Using items generated through rigorous consensus methodology and applying modern validity theory as a framework, our research team developed and tested the eSQUARE, a novel 10-item assessment tool, to assess the quality of eConsult replies.

Our study provides sources of validity evidence for the eSQUARE, including content, response process and internal structure evidence. Having a group of medical education researchers with expertise in assessment tool development and psychometric analyses,

plus a rigorous method of selecting items to ensure that eSQUARE elements represent the quality of eConsult specialist written communication, demonstrates content evidence.³⁶ We infer that rater training was adequate, since participants did not ask for any clarification nor were any revisions required for the instructional PDF document, thus generating response process validity.^{37,38} The eSQUARE was able to discriminate between eConsults rated as low, medium or high quality as defined in Table 1; having raters consistently rate eConsults in a similar matter builds further response process evidence. Our analyses found high reliability measures, including a G-coefficient of 0.96, indicating the eSQUARE could produce consistent scores and thus demonstrating internal structure evidence.^{21,39}

The eSQUARE performed well when compared to other written communication instruments. While we ultimately recruited only 8 out of our target of 10 raters, our D-study demonstrated that just a single rater assessing one eConsult was adequate to achieve a reliability measure of ≥ 0.70 with our sample of eConsults, a described benchmark for reliable formative assessment.³⁵ This compares favourably to the Sheffield Assessment Instrument for Letters (SAIL),⁴⁰ a tool assessing the quality of clinic letters written by specialist registrars, where three raters applying the

SAIL to eight letters for a given registrar are required to achieve a reliability measure of 0.70. This has future implications, since requiring multiple raters and eConsults per specialist could impair the feasibility of widespread implementation of the eSQUARE. Our figures also align with literature describing convenience samples of clinical performance, where 7–11 observer ratings are needed to produce adequate generalisability data, for example a G-coefficient ≥ 0.80 as seen in our study.⁴¹

The eSQUARE can potentially be used to guide new specialists on eConsult services, as well as to provide feedback/quality control. Its ease of use and high reliability, limiting the number of raters required, make it well-suited to wide applicability. Our raters did not ask for any further clarification on how to use the eSQUARE, nor did they seek training beyond the single instructional document provided to them. This compares favourably to rater training described for other medical documentation assessment tools that describe one- to two-hour training sessions with the primary investigator.^{24,42} We thus do not feel that formal training is required to use the eSQUARE, which should facilitate its widespread implementation.⁴³ Our raters used the entire range of the five-point scales for each item (Table 2). Thus, there was no evidence of end-aversion or central-tendency biases where raters tend to avoid low scores or favour middle options.⁴⁴

Our analysis included item total correlations, inter-item correlations and a generalisability study which demonstrated very high reliability measures. As these may indicate item redundancy, one could argue that the eSQUARE could be condensed further. For example, one could combine the items ‘Anticipatory guidance’ and ‘When face-to-face referral needed’, since both had the highest number of N/A ratings, the latter item had the lowest reliability measures and their inter-item correlation was highest compared to other items. These two items could be combined as:

Anticipatory guidance including red flags (e.g. key features that would prompt further work-up), what to try next if recommendations do not result in a favourable outcome, and **when a face-to-face referral would be indicated.**

Since we intend for the eSQUARE to be formative assessment tool, we opted not to combine items further, despite high reliability measures. The remaining items represent those generated by formal consensus methods. Thus, each item may provide its own value when providing feedback to specialists while still achieving reasonable psychometric rigour for the eSQUARE.

Study limitations

This was a single-centre Canadian study that only recruited users of the Champlain BASE™ eConsult programme and may not be generalisable to other eConsult platforms and providers. For example, eConsult specialists who share common electronic records with the PCP – a rare occurrence within our service – have resources to find information not originally presented by the PCP. Additional limitations may be exposed upon implementation of eSQUARE on a broader scale. Since we tested the eSQUARE using handpicked eConsults on a limited number of specialties, reliability measures for a random eConsult sample may be less robust. Also, since raters may have volunteered to participate in the study due to personal interest in eConsult assessment, ‘real-world’ raters who are not as engaged as study participants may produce lower levels of reliability when using the eSQUARE. While our survey data indicate that most eConsults are highly rated by PCPs, this may be driven by elements not captured by eSQUARE such as timeliness of response and reluctance to rate colleagues harshly.

Future directions

By using items that describe features of high-quality eConsults, the eSQUARE can inform educational interventions aiming to improve specialist-to-PCP communication via eConsult. One approach is audit and feedback⁴⁵ where eSQUARE scores can be combined with other outcome data (e.g. PCP survey data, referral outcomes for patients) to measure how a single specialist or group of consulting physicians within a specialty are performing with respect to their peers. This information could help identify underperforming specialists for targeted education and could add to performance reports for the purpose of satisfying accreditation bodies.⁴⁶ The eSQUARE tool can be included in eConsult educational activities for faculty or resident physicians (e.g. accredited workshops, seminars, short courses). eSQUARE scores can be used as outcome measures when assessing the impact of these educational interventions, similar to how other written communication assessment tools have demonstrated improvements in rating scores^{17,47} and favourable changes in letter-writing behaviours.¹⁸

Conclusions

Using modern validity theory as a framework, we have demonstrated initial validity evidence, including multiple reliability measures, for the novel eSQUARE tool. Since a single rater without extensive training can achieve adequate reliability measures for the purpose of formative feedback, the eSQUARE is a feasible tool

for the assessment of the quality of specialist-to-PCP communication via eConsult. Our next steps include applying the eSQUARE to a larger sample of eConsults to build further validity evidence and incorporating the eSQUARE tool into educational interventions that aim to improve eConsult correspondence, both within our regional system and abroad.

Acknowledgements

We thank Sheena Guglani, Katherine Scowcroft and Justin Joschko for their assistance.


Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship and/or publication of this article: Department of Medicine Medical Education Research Grant, University of Ottawa, Canada.

ORCID iD

Christopher Tran  <https://orcid.org/0000-0001-6708-4884>

References

- O'Malley A, Reschovsky J and Saiontz-Martinez C. Interspecialty communication supported by health information technology associated with lower hospitalization rates for ambulatory care-sensitive conditions. *J Am Board Fam Med* 2015; 28: 404–417.
- Vermeir P, Vandijck D, Degroote S, et al. Communication in healthcare: a narrative review of the literature and practical recommendations. *Int J Clin Pract* 2015; 69: 1257–1267.
- O'Malley A and Reschovsky J. Referral and consultation communication between primary care and specialist physicians. *Arch Intern Med* 2011; 171: 56–65.
- Vimalananda V, Gupte G, Seraj S, et al. Electronic consultations (e-consults) to improve access to specialty care: a systematic review and narrative synthesis. *J Telemed Telecare* 2015; 21: 323–330.
- Olayiwola J, Potapov A, Gordon A, et al. Electronic consultation impact from the primary care clinician perspective: outcomes from a national sample. *J Telemed Telecare* 2019; 25: 493–498.
- Kirsh S, Carey E, Aron D, et al. Impact of a national specialty e-consultation implementation project on access. *Am J Manag Care* 2015; 21: e648–654.
- Tuot D, Murphy E, McCulloch C, et al. Leveraging an electronic referral system to build a medical neighborhood. *Healthcare* 2015; 3: 202–208.
- Joschko J, Keely E, Grant R, et al. Electronic consultation services worldwide: environmental scan. *J Med Internet Res* 2018; 20: e11112.
- Keely E and Liddy C. Transforming the specialist referral and consultation process in Canada. *Can Med Assoc J* 2019; 191: E408–409.
- Keely E, Drosinis P, Afkham A, et al. Perspectives of Champlain BASE specialist physicians: their motivation, experiences and recommendations for providing eConsultations to primary care providers. *Stud Health Technol Inform* 2015; 209: 38–45.
- Keely E, Williams R, Epstein G, et al. Specialist perspectives on Ontario provincial electronic consultation services. *Telemed J E Health* 2019; 25: 3–10.
- Liddy C, Afkham A, Drosinis P, et al. Impact of and satisfaction with a new eConsult service: a mixed methods study of primary care providers. *J Am Board Fam Med* 2015; 28: 394–403.
- Liddy C, Moroz I, Mihan A, et al. A systematic review of asynchronous, provider-to-provider, electronic consultation services to improve access to specialty care available worldwide. *Telemed J E Health* 2019; 25: 184–198.
- Lee M, Ray K, Mehrotra A, et al. Primary care practitioners' perceptions of electronic consult systems. *JAMA Intern Med* 2018; 178: 782–789.
- Rash A, Sheldon R, Donald M, et al. Valued components of a consultant letter from referring physicians' perspective: a systematic literature synthesis. *J Gen Intern Med* 2018; 33: 948–954.
- Russell C, Sandu V, Moroz I, et al. Key components of traditional consultation letters and their relevance to electronic consultation replies: a systematic review. *Telemed J E Health* 2020; 26: 689–699.
- Fox A, Palmer R, Crossley K, et al. Improving the quality of outpatient clinic letters using the Sheffield Assessment Instrument for Letters (SAIL). *Med Educ* 2004; 38: 852–858.
- Keely E, Myers K, Dojeiji S, et al. Peer assessment of outpatient consultation letters – feasibility and satisfaction. *BMC Med Educ* 2007; 7: 13.
- Tuot D, Sehgal N, Neeman N, et al. Enhancing quality of trainee-written consultation notes. *Am J Med* 2012; 125: 649–652.
- Tran C, Archibald D, Humphrey-Murto S, et al. What makes a high-quality electronic consultation (eConsult)? A nominal group study. *J Telemed Telecare* 2020; 26: 239–247.
- Cook D and Beckman T. Current concepts in validity and reliability for psychometric instruments: theory and application. *Am J Med* 2006; 119: 166.e7–16.
- Halman S, Dudek N, Wood T, et al. Direct observation of clinical skills feedback scale: development and validity evidence. *Teach Learn Med* 2016; 28: 385–394.
- Bierman J, Hufmeyer K, Liss D, et al. Promoting responsible electronic documentation: validity evidence for a checklist to assess progress notes in the electronic health record. *Teach Learn Med* 2017; 29: 420–432.
- Weber D, Held J, Jandarov R, et al. Development and establishment of initial validity evidence for a novel tool for assessing trainee admission notes. *J Gen Intern Med* 2020; 35: 1078–1083.
- Stansfield R and Gruppen L. Power analyses: planning, conducting and evaluating education research. In: J

- Cleland and SJ Durning (eds) *Research medical education*. Oxford: John Wiley, 2015, pp.43–50.
26. Hertzog M. Considerations in determining sample size for pilot studies. *Res Nurs Health* 2008; 31: 180–191.
 27. Johanson G and Brooks G. Initial scale development: sample size for pilot studies. *Educ Psychol Meas* 2010; 70: 394–400.
 28. Crossley J, Russell J, Jolly B, et al. ‘I’m pickin’ up good regressions’: the governance of generalisability analyses. *Med Educ* 2007; 41: 926–934.
 29. McMaster University. The program for educational research and development, https://healthsci.mcmaster.ca/merit/research/g_string_v (accessed 2 October 2020).
 30. Brennan R. GENOVA suite programs, <https://education.uiowa.edu/centers/center-advanced-studies-measurement-and-assessment/computer-programs#GENOVA> (accessed 2 October 2020).
 31. Bloch R and Norman G. G_String V: user manual, https://healthsci.mcmaster.ca/docs/librariesprovider15/default-document-library/g-string-v5-user-manual.pdf?sfvrsn=9c6f7395_2 (accessed 2 October 2020).
 32. Vispoel W, Morris C and Kilinc M. Practical applications of generalizability theory for designing, evaluating, and improving psychological assessments. *J Pers Assess* 2018; 100: 53–67.
 33. Norman G. Likert scales, levels of measurement and the ‘laws’ of statistics. *Adv Health Sci Educ* 2010; 15: 625–632.
 34. Willits F, Theodori G and Luloff A. Another look at Likert scales. *J Rural Soc Sci* 2016; 31: 126–139.
 35. Downing S. Reliability: on the reproducibility of assessment data. *Med Educ* 2004; 38: 1006–1012.
 36. Edwards M, Slagle A, Rubright J, et al. Fit for purpose and modern validity theory in clinical outcomes assessment. *Qual Life Res* 2018; 27: 1711–1720.
 37. Lineberry M. *Validity and quality. Assessment in health professions education*. New York: Routledge, 2019.
 38. Reeves T and Marbach-Ad G. Contemporary test validity in theory and practice: a primer for discipline-based education researchers. *CBE Life Sci Educ* 2016; 15: 1–9.
 39. Park Y. *Reliability. Assessment in health professions education*. New York: Routledge, 2019.
 40. Crossley J, Howe A, Newble D, et al. Sheffield Assessment Instrument for Letters (SAIL): performance assessment using outpatient letters. *Med Educ* 2001; 35: 1115–1124.
 41. Williams R, Klamen D and McGaghie W. Cognitive, social and environmental sources of bias in clinical performance ratings. *Teach Learn Med* 2003; 15: 270–292.
 42. Stille C, Mazor K, Meterko V, et al. Development and validation of a tool to improve paediatric referral/consultation communication. *BMJ Qual Saf* 2011; 20: 692–697.
 43. Cook D, Dupras D, Beckman T, et al. Effect of rater training on reliability and accuracy of mini-CEX scores: a randomized, controlled trial. *J Gen Intern Med* 2009; 24: 74–79.
 44. Streiner D, Norman G and Cairney J. *Biases in responding. Health measurement scales*. Oxford: Oxford University Press, 2015, pp.100–130.
 45. Colquhoun H, Brehaut J, Sales A, et al. A systematic review of the use of theory in randomized controlled trials of audit and feedback. *Implement Sci* 2013; 8: 66.
 46. Keely E, Archibald D, Tuot D, et al. Unique educational opportunities for PCPs and specialists arising from electronic consultation services. *Acad Med* 2017; 92: 45–51.
 47. Sewell J, Day L, Tuot D, et al. A brief, low-cost intervention improves the quality of ambulatory gastroenterology consultation notes. *Am J Med* 2013; 126: 732–738.

Appendix I: Explanations of individual eSQUARE items provided to raters

Item Guide

Remember: try to assign a value for each item.

Use ‘N/A’ if item absolutely does not apply to the case.

1. Is **current (e.g. up to date)** and evidence based where applicable.

Nominal group participants felt eConsult advice should be based on latest evidence if possible. We recognise that you may not be a content expert; please use your best judgement for this item.

2. Is **helpful and/or educational**, e.g. the specialist **provides rationale** or evidence for recommendations, or attaches additional material such as guidelines or clinical pathways.

Both primary care providers and specialists recognise the potential educational value of eConsult. A high-quality eConsult would be one where the specialist does more than just answer the question, but also provides some education around the case for the recipient to apply to future cases. This is often in the form of rationale for their recommendations, i.e. ‘This is why I am making these recommendations’.

3. Is **patient specific** by incorporating available patient information to the clinical question(s).

While some clinical questions are frequently asked via eConsult, nominal group participants did not want the specialist to provide generic, ‘copy-and-paste’ advice for a common clinical scenario. The nuances of a particular case and how they specifically relate to each individual patient should be acknowledged. A high-quality eConsult provides advice that is catered to the specific patient in question.

4. Addresses each question/concern posed by the referring provider.

A high-quality eConsult not only answers each question, but also acknowledges any concerns expressed by the referring provider that may not be specifically posed as a question. For example, in response to 'What medication would you start? I am worried about side effects', the specialist has two things to address.

5. **Specific recommendations**, including cost and availability of tests and treatment, while acknowledging if suggestions are off label. Medication advice includes how to dose, titrate and monitor response to therapy.

Nominal group participants appreciated a 'don't hold back' approach where detailed information for each proposed test or treatment are given. For example, a lower-quality eConsult might simply say 'start metformin and increase dose as tolerated, and check HbA1C in three months', whereas a higher-quality eConsult may mention 'I recommend metformin 500 mg b.i.d. × 1 week, then up-titrate to 1000 mg b.i.d. as tolerated (GI side effects). It is relatively inexpensive and readily available. You can expect a 1–1.5% HbA1C reduction in three months'.

6. **Anticipatory guidance** including red flags (e.g. key features that would prompt further work-up) and what to try next if recommendations do not result in a favourable outcome.

Nominal group participants wanted more than just one plan of action. A high-quality eConsult would provide an order of suggestions to follow, particularly in 'red flag' situations. For example, 'If treatment A does not work, try treatment B, followed by C if you are not quite at target. If treatment A causes agranulocytosis, then stop the medication and refer that patient to emergency'.

7. Specific advice as to when a face-to-face referral would be indicated.

More than 60% of referring providers contemplate a face-to-face referral before submitting an eConsult. A

high-quality eConsult response will outline specific advice to when the eConsult should be converted to a traditional request for face-to-face consultation. For example, 'Try treatments A and B first. If they do not work, then refer to my clinic to help arrange endoscopy'.

8. **Action items that are doable** with references to local resources where appropriate.

Since the referring provider nearly always remains the most responsible for carrying out specialist advice, a high-quality eConsult provides advice that the referring provider can actually implement, such as common laboratory or imaging studies accessible to the referring provider. Examples of not readily implementable advice include procedures requiring specialist expertise (e.g. intra-articular platelet-rich plasma injections) or parental therapeutics (e.g. intravenous iron or bisphosphonates).

9. **Clear and organised** using logical flow where key information is easy to find.

A high-quality eConsult flows well and is easy to follow.

10. Delivered in a **professional, supportive tone** (e.g. specialist acknowledges a difficult and/or complex case) that is open to back-and-forth communication.

Nominal group participants appreciated having collegiality between health-care providers. A high-quality eConsult gives off a clear impression that the specialist is trying to understand where the referring provider is coming from, and leaves it open for the referring provider to ask further questions.

11. How would you rate the overall quality of specialist's advice?

Compare the eConsult to the responses you have read in this study and also the ones you may have read in your own experience. Where does the eConsult rate in terms of overall quality?

Appendix 2: eSQUARE decision study

		Number of raters/eSQUARE							
		1	2	3	4	5	6	7	8
Number of scale items	1	0.62	0.77	0.83	0.87	0.89	0.91	0.92	0.93
	2	0.68	0.81	0.86	0.89	0.91	0.93	0.94	0.94
	3	0.70	0.82	0.88	0.90	0.92	0.93	0.94	0.95
	4	0.71	0.83	0.88	0.91	0.93	0.94	0.95	0.95
	5	0.72	0.84	0.89	0.91	0.93	0.94	0.95	0.95
	6	0.73	0.84	0.89	0.91	0.93	0.94	0.95	0.96
	7	0.73	0.84	0.89	0.92	0.93	0.94	0.95	0.96
	8	0.73	0.85	0.89	0.92	0.93	0.94	0.95	0.96
	9	0.74	0.85	0.89	0.92	0.93	0.94	0.95	0.96
	10	0.74	0.85	0.89	0.92	0.93	0.94	0.95	0.96