American Journal of Pharmaceutical Education 2018; 82 (9) Article 6725.

EDUCATION BRIEF

Development and Validation of a Rubric to Evaluate Diabetes SOAP Note Writing in APPE

Miranda R. Andrus, PharmD,[a] Sharon L. K. McDonough, PhD,[b] Kristi W. Kelley, PharmD,[a] Pamela L. Stamm, PharmD,[a] Emily K. McCoy, PharmD,[a] Katelin M. Lisenby, PharmD,[a] Heather P. Whitley, PharmD,[a] Nicole Slater, PharmD,[a] Dana G. Carroll, PharmD,[a] E. Kelly Hester, PharmD,[a] Allison Meyer Helmer, PharmD,[a] Cherry W. Jackson, PharmD,[a] Debbie C. Byrd, PharmD, MBA[c]

[a] Harrison School of Pharmacy, Auburn University, Auburn, Alabama
[b] University of Tennessee Health Science Center College of Pharmacy, Memphis, Tennessee
[c] Bill Gatton College of Pharmacy, East Tennessee State University

Submitted August 16, 2017; accepted January 7, 2018; published November 2018.

Objective. To develop and establish validity for a grading rubric to evaluate diabetes subjective, objective, assessment, plan (SOAP) note writing on primary care (PC) advanced pharmacy practice experiences (APPEs), and to assess reliability and student perceptions of the rubric.

Methods. Ten PC APPE faculty members collaborated to develop a rubric to provide formative and summative feedback on three written SOAP notes per APPE student over a 10-month period. Correlation analyses were conducted between rubric scores and three criterion variables to assess criterion-related validity: APPE grades, Pharmaceutical Care Ability Profile Scores, and Global Impression Scores. Inter-rater and intra-rater reliability testing were completed using Cohen's kappa and Intraclass Correlation Coefficients (ICC). Student perceptions were assessed through an anonymous student survey.

Results. Fifty-one students and 167 SOAP notes were evaluated using the final rubric. The mean score significantly increased from the first to second SOAP note and from the first to third SOAP note. Statistically significant positive correlations were found between final rubric scores and criterion variables. The ICC for inter-rater reliability was fair (.59) for final rubric scores and excellent for intra-rater reliability (.98 to1.00). Students responded that the rubric improved their ability (84.9%) and confidence (92.4%) to write SOAP notes.

Conclusion. The rubric may be used to make valid decisions about students' SOAP note writing ability and may increase their confidence in this area. The use of the rubric allows for greater reliability among multiple graders, supporting grading consistency.

Keywords: SOAP note, rubric, experiential education, reliability, validity

Downloaded from http://www.ajpe.org by guest on April 26, 2021. © 2018 American Association of Colleges of Pharmacy

INTRODUCTION

Communication through clear, accurate, succinct, and consistent documentation is an essential component to providing safe and effective medication therapy within the medication use system.[1,2] In addition to demonstrating value within the health care system, effective documentation affords pharmacists the ability to communicate patient care recommendations with other health care providers, increases awareness of the pharmacist's role in provision of safe and effective medication therapy management, and holds pharmacists professionally responsible for recommendations.[2-4] Professional documentation is commonly written using the structured format of the SOAP (subjective, objective, assessment, plan) note. Using this approach encourages a specific clinical thought process and requires students to use critical thinking and problem-solving skills.[5,6]

A recent assessment of SOAP note evaluation tools examined assessment tools for SOAP notes used by laboratory instructors at pharmacy colleges and schools.[7] There was a wide degree of variation in the tools and their content. The authors noted that the lack of consistency in assessment tools may lead to variation in documentation

Corresponding Author: Sharon L. K. McDonough. Tel: 334-524-7213. E-mail: sharonleemcdon@gmail.com

Note: At the time of this research, Sharon McDonough was coordinator of assessment and an assistant professor at the University of Tennessee Health Science Center College of Pharmacy, Memphis, Tennessee.

Downloaded from http://www.ajpe.org by guest on April 26, 2021. © 2018 American Association of Colleges of Pharmacy

of patient care when students enter practice. To date, studies evaluating student pharmacists' clinical documentation have been conducted only in the early portion of curricula.[4-6,8,9] Student pharmacists are expected to have basic skills to effectively document patient care activities when entering advanced pharmacy practice experiences (APPE), yet a paucity of data exists evaluating the quality of student documentation during direct patient care experiences. Despite these challenges, students perceive the use of grading rubrics to be fair and transparent.[10]

The Accreditation Council for Pharmacy Education (ACPE) emphasizes the need for systematic valid, reliable assessment tools, both for organizational purposes and student learning outcomes (Standard 24).[11] Thus, incorporation of a formative assessment of written SOAP notes using a standardized grading rubric could augment evaluation of student performance in APPEs.

At the Auburn University Harrison School of Pharmacy, a traditional, four-professional-year school, students are initially taught SOAP note writing in the fall semester of the first professional (P1) year in a clinical skills laboratory course. Students then apply these skills during their Introductory Pharmacy Practice Experience (IPPE) patient care documentation throughout the first through third professional (P3) years. Additionally, P3 students are expected to write various components of SOAP notes in their required pharmacotherapy course, a clinical skills laboratory course, and electives. Throughout the fourth professional (P4) year, students are evaluated on professional writing in the summative experiential course evaluation; however, no specific evaluation tool is available for SOAP note writing. Both students and faculty acknowledge the need for a consistent approach for SOAP note writing in APPE. The objective of this study was to collaboratively develop and establish the validity of a grading rubric for evaluation of diabetes SOAP notes on primary care APPEs. The secondary objectives were to assess inter-rater reliability, intra-rater reliability, and student perceptions of the rubric.

METHODS

In April 2015, a cohort of 10 primary care APPE full-time faculty at Auburn University responded to an open invitation to develop an evaluation instrument for SOAP notes. After an extensive literature search, faculty selected the process suggested by Mertler and colleagues and Allen and colleagues for development and validation of the rubric.[12,13] The following objectives were established for the project: provide clarification for students on how to write a "real life" SOAP note, provide feedback to improve SOAP note writing, be transferable to other preceptor sites and disease states, and be practical and efficient (goal 10-15 minutes for feedback). Faculty discussed current processes for SOAP note assignments, obtained example rubrics from faculty, and conducted a literature review. The group discussed a general vs a disease-state specific rubric and common types of patients at all practices. The group decided to start with a rubric specific to diabetes visits in order to provide concrete examples, with plans to generalize it in the future. Criteria were designed in a way that could be applied and assessed in other types of patients. A small working group developed a draft analytic rubric identifying four domains (subjective, objective, assessment, and plan), criteria for each domain with three rating scales (did not perform, performed incompletely, and performed), and specific performance descriptors with examples for many criteria to address reliability. A "general" domain was also included to evaluate overall appearance, organization, and appropriate placement of data into the S, O, A and P sections. The first version was distributed and was approved for use by the faculty in May 2015. The faculty did not complete formal training on how to use the rubric. Instead, a consensus building approach was used as all faculty involved served as experts in SOAP note writing, and training was conducted informally as the rubric was used over time. Each faculty cohort member provided the rubric in orientation and required students to submit three SOAP notes for evaluation over the five-week APPE. The first SOAP was written during the first two weeks, and the completed rubric was provided as formative feedback only. The second and third SOAP notes were submitted over the remaining three weeks of the APPE. These scores were averaged and contributed 5% to10% to the overall APPE grade.

During the first three APPE blocks, cohort members discussed the rubric continually via an email thread to determine functionality and to ensure grading consistency and improve reliability. The rubric underwent two revisions based on the experiences of faculty members. Revisions included weighing of domains in the calculation of the grade (ie, assessment score > plan score > subjective score), combining repetitive sections, modifying descriptors to be more specific, and providing more examples to decrease the need for written comments. Ratings were reduced to yes/no (complete/incomplete) for all of the objective measures, some of the subjective measures (ie, demographics), non-pharmacologic assessment, plans for follow-up, and all general domain descriptors. These sections were combined in the grading and referred to as a combined yes/no score. A Global Impression Score (GIS) was also collected prior to completion of the rubric (the faculty's assessment score reflecting the letter grade they would have expected to

Downloaded from http://www.ajpe.org by guest on April 26, 2021. © 2018 American Association of Colleges of Pharmacy

assign the SOAP note). The final version of the rubric was implemented in September 2015 and is available upon request. The rubric was primarily used for diabetes SOAP notes, but occasionally used for pre-diabetes or non-diabetes notes so that students could complete the assignments if patients with diabetes were not available.

The researchers' university institutional review boards approved this study. Rubrics completed for students of the faculty cohort during May 2015-February 2016 were included for evaluation. Data from faculty evaluations of SOAP notes were collected via standardized forms using Qualtrics (Qualtrics LLC, Provo, UT). These data included history of previous SOAP writing in primary care, previous evaluations via the rubric, type of SOAP note (diabetes, pre-diabetes, or non-diabetes), time it took to complete the evaluation, the evaluator's GIS, the rubric score for each criterion, and the overall scores for each domain: subjective, assessment, plan and combined yes/no scores. Data were imported from Qualtrics into Statistical Package for Social Sciences (SPSS) Version 22.0 (Chicago, IL) for statistical analyses.

Correlation analyses were conducted to assess criterion-related validity. Pearson correlation coefficients were calculated to examine correlations between final rubric scores and the following three criterion variables: GIS, Primary Care APPE overall grades, and Pharmaceutical Care Ability Profile (PCAP) score. The PCAP is an APPE-validated assessment rubric previously developed by a consortium of schools and used to calculate at least 60% of a student's APPE grade.

Inter-rater reliability was assessed using a sample of SOAP notes that were evaluated by all 10 cohort members. One investigator used coded lists of faculty and students to randomly select six notes from those graded with the final rubric, ensuring each note was from a different practice site and faculty member. The investigator collected and stored de-identified notes on a centralized, secured server. Cohort members then graded each of the six notes for testing of intra-rater and inter-rater reliability of the rubric. Raters were blinded to the note author, written APPE week (1 to 5) or date note, APPE preceptor, and note number (1 to 3). Intraclass Correlation Coefficients (ICC) were calculated to assess consistency among the 10 raters on each of the subscores and on the final numeric score. A two-way random model, consistency and average measures were used in the ICC calculation process. Inter-rater consistency for individual items of the rubric was tested using a variation of Cohen's kappa developed by Light for estimating consistency among multiple raters.[14] This adjusted statistic represents the average kappa for all raters on a given item by calculating Cohen's kappa for each pair of raters and then calculating the arithmetic mean for all pairs (n=45 pairs per item). Intra-rater reliability (test-retest) was assessed for the final rubric by six cohort members using original and repeat evaluations of a single SOAP note. Due to risk of bias, the six repeat graders were instructed to avoid reviewing the previous scores or feedback provided prior to regrading the notes. Graded rubric scores were entered into a Qualtrics survey for data collection. Pearson's r correlations were then calculated for each section of the rubric and the final score to assess consistency in rating the same SOAP note on two different occasions. The a priori level of statistical significance was established as $p \leq .05$ for all analyses.

From May 2015 to April 2016, 95 students who completed their primary care APPE with the faculty cohort were asked to participate in a survey to collect information on their experience using the SOAP note rubric. Students completing multiple primary care APPEs were eligible to complete the survey only during their first primary care APPE with the faculty cohort. The survey was designed to gather student perceptions on the following areas: whether the prior pharmacy curriculum correlated to expectations of the rubric, if the rubric improved their ability to write a SOAP note, if the rubric adequately assessed their ability to write a SOAP note, if the rubric was fair and graded them appropriately, and whether or not they received constructive feedback using the rubric to improve future notes. Surveys were distributed via email to students at the end of each APPE via Qualtrics for data collection. All survey responses were anonymous, and participation was voluntary.

RESULTS

From May 2015 to February 2016, there were 79 students and 276 SOAP notes evaluated, and eight of the 79 students had SOAP notes evaluated during different blocks using two different versions of the rubric. Only data from the final version of the rubric are presented (51 students and 167 SOAP notes) (Table 1). Four SOAP notes were excluded from the final version's data set because multiple answers were chosen ("half credit") during data entry. The mean final PCAP and overall APPE grades ranged from a mid- to high-B letter grade.

The SOAP note scores for the final version of the rubric are detailed in Table 2. Final percent rubric scores consistently increased from the first SOAP to the third SOAP with the lowest scores on the assessment section (weighted 40%). Using the final rubric, the mean SOAP note score significantly increased from the first to second SOAP note ($p<.001$) and from the first to third SOAP note ($p<.001$). This was also true for the first and second versions of the rubric. There were no differences in scores

Downloaded from http://www.ajpe.org by guest on April 26, 2021. © 2018 American Association of Colleges of Pharmacy

Table 1. SOAP Notes Evaluated Using a Standard Rubric on Primary Care APPEs

Characteristic	Final Rubric (n=51 P4s)
SOAP number, N (%)	
One	59 (35.3)
Two	56 (33.5)
Three	52 (31.1)
SOAP type, N (%)	
Diabetes	63 (90.0)
Prediabetes	1 (1.4)
Non-diabetes	6 (8.6)
Third SOAP grade, Mean (SD)	91.5 (7.7)
Third SOAP Global Impression Score (GIS)[a] Mean (SD)	4.55 (0.6)
Final PCAP grade, Mean (SD)	89.3 (7.1)
APPE grade, Mean (SD)	89.0 (5.9)

[a] Letter grades were converted to a 5-point scale for statistical analysis

between students who had completed a previous APPE in primary care with SOAP note writing compared to those who had not (p=.19). There were also no differences in scores based on the APPE block in which they completed the note (p=.25). Preceptors reported a Mean (SD) time of 23.0 (0.9) minutes to grade a single SOAP note, with the time decreasing by approximately 5 minutes from the first to the third note.

Statistically significant positive correlations were found between final rubric scores and all three criterion variables for all three versions of the rubric. The strongest correlations were between the final percentage score and GISs for the final version of the rubric. Regarding relationships between individual sections of the rubric (subjective, objective, assessment, combined yes/no) and the criterion variables (ie, APPE Grade, PCAP Grade, GIS), significant positive correlations were found for all comparisons.

ICC values ranged from poor to good according to Cicchetti's interpretation,[15] with the subjective raw score having the lowest performance at .28. The combined yes/no percent score section classified as good reliability at .69. The assessment raw score, plan raw score, and the final percent score all had fair reliability at .49, .56, and .59, respectively.

Cohen's kappa for inter-rater consistency results ranged from -.05 for Professional Appearance (in the General Domain) to .77 for Vital Signs (in the Objective Domain). It should be noted that Cohen's kappa is not calculated when one of the variables is a constant (ie, the rater's answers did not vary on the item across the six SOAP notes). When this was the case, only the number of pairs for which kappa could be calculated were included in the final mean kappa value. One item, Allergies (in the Subjective Domain), was excluded from this analysis because all but two variables in the pairs were constants.

Six randomly selected notes were rated at two different times (eight to 13 months apart) by six cohort members to examine test-retest reliability. Results showed strong positive correlations in which three of the five measures were statistically significant (Table 3). Additionally, ICC were calculated for each rater's final scores. All ICC values showed excellent consistency, ranging from .98 to 1.00.

Fifty-three of the 95 students who completed APPEs from May 2015 to April 2016 completed the survey, resulting in a 55.8% response rate. A majority (88.7%) of students responded that they had a clear understanding of expectations for SOAP notes after reviewing the rubric, and 84.9% felt the rubric improved their ability to write a SOAP note. Roughly 74% indicated the rubric helped to provide an adequate assessment of their notes, and 92.5% indicated they received constructive feedback from their preceptor(s) to improve future notes. While only half (50.9%) of students were confident in their SOAP note writing ability prior to the start of their fourth year APPEs, 92.4% indicated they were more confident after completing their primary care APPE.

Table 2. SOAP Note Scores Using the Final Rubric

SOAP Section	SOAP 1 (n=59)	SOAP 2 (n=56)	SOAP 3 (n=52)
Yes/No[a], M %	80.2	91.3	91.9
Subjective, M %	81.0	93.0	94.0
Assessment, M %	65.9	84.4	89.6
Plan score, M %	73.1	86.1	91.1
Final score, M % (SD)	73.5 (16.3)	88.2 (12.5)	91.5 (7.7)

[a] Yes/No combined score includes some S, A and P data, all of the O data and general evaluation of writing
M=mean, SD=standard deviation

DISCUSSION

In this study, fourth professional year student pharmacists displayed an increased ability to document patient care activities in the SOAP note format with repeated practice and evaluation with a standardized rubric developed by primary care faculty. It is possible that improvement was due to improved understanding of APPE expectations and knowledge of disease state goals. Regardless of the rubric version used, a statistically significant improvement in scores was seen from the first to the third SOAP note. However, there was no significant difference found in scores between students who had

Downloaded from http://www.ajpe.org by guest on April 26, 2021. © 2018 American Association of Colleges of Pharmacy

Table 3. Pearson's Correlations Between Time 1 and Time 2 for Test-Retest Reliability

	Pearson's *r*	*p* value
Global impression score	.75	.086
Combined Yes/No percent score[a]	.95	.003[c]
Subjective raw score	.91	.012[b]
Assessment raw score	.66	.16
Final percent score	.86	.028[b]

[a] Yes/No combined score includes some S, A and P data, all of the O data and general evaluation of writing
[b] Correlation is significant at the .05 level
[c] Correlation is significant at the .01 level

completed SOAP notes during a previous primary care APPE compared to those who had not; this could be attributed to the differences in overall SOAP note writing expectations amongst preceptors. This result aligns with Shogbon and colleagues and Brown and colleagues' findings.[4,5] Shogbon and colleagues reported incorporation of a faculty-developed objective grading rubric in a second-year therapeutics course.[4] The students' overall ability significantly improved from the first to the fifth patient care note (85.8% vs 91.8%; $p<.001$) despite cumulative content material and increasing difficulty of topics assessed. Brown and colleagues' incorporation of the SOAP note format during a third-year pharmacotherapy course showed that overall mean SOAP note scores improved from 3.23 to 3.62 (out of 4) from the first to the fourth note ($p<.001$).[5]

Students consistently struggle with proper assessment of patient care goals. On each SOAP note, the lowest scores seen were in the assessment section, which is consistent with other studies. Brown and colleagues observed that lower quality assessment and plan sections were seen on the first note (54% needed improvement), although that improved with each successive attempt (27% needed improvement by the fourth note, $p<.001$), and errors in assessment occurred twice as often as in the plan section (20.5% vs 10.5% of notes, respectively) of the SOAP note.[5] Other studies have also noted that assessment documentation is more often incomplete vs other sections of the note.[7,16] The findings of our study align with early curriculum studies in which students consistently struggled with the assessment sections, but that repeated SOAP note writing opportunities improved students' confidence and abilities.[4-6,16]

Grading rubrics are not only useful tools to evaluate student performance and provide formative assessments; studies have shown that students perceive the use of rubrics to be fair and transparent.[10] In this study, incorporation of a SOAP note rubric in APPEs improved student understanding of expectations for documentation and positively affected student perception of SOAP note assessment and formative feedback by the preceptor. Additionally, student confidence in their SOAP note writing ability improved approximately 40% by the end of the APPE. While the study goal of 10-15 minutes per note for feedback was not met, the time required for feedback decreased as students gained more SOAP note writing experiences.

Rubric reliability has been found to increase with repeated use.[10] Each preceptor used this SOAP note rubric three times per student evaluated. Although some individual items on the rubric did not show strong consistency in the inter-rater reliability analysis, the overall subscores showed very good inter-rater reliability. The analyses suggest that more work may be needed to ensure that raters interpret all criteria consistently.

There are some statistical limitations to this study. One is the small sample size of SOAP notes used to test inter- and intra-rater reliability, which was determined for feasibility given the necessary grading time. Another is the inability to compute Cohen's kappa when one of the variables is a constant, thereby reducing the number of pairs that could be included in the mean calculation of kappa. An option might have been to compute a simple percentage agreement between pairs and average the percentages to arrive at a mean percentage agreement. However, the decision was made not to do so, given that Cohen and Krippendorff have rejected the percentage agreement procedure entirely as an adequate inter-rater reliability measure because it inflates agreement estimates by not taking into account chance agreement.[17,18] Arguably, the use of Light's procedure to adjust kappa for multiple raters, which was used in this study, may actually have deflated inter-rater agreement estimates by excluding some pair comparisons from the analysis.[14]

Another limitation is the generalizability of the rubric. The instrument was not tested with affiliate preceptors due to time and feasibility constraints. This rubric was developed specifically for diabetes-focused SOAP notes, but the investigators were able to adapt it for use on SOAP notes written for pre-diabetes and non-diabetic patients. The rubric has since been modified for non-disease specific (or general) SOAP notes. While the survey responses indicated an overall positive perception from students, the low response rate could have affected the results of the survey.

CONCLUSION

The rubric developed and validated in this study provides students with clear expectations of SOAP note writing in primary care APPEs, significantly improves their ability to write SOAP notes, and may contribute to

increased student confidence in writing SOAP notes. Validity evidence for use of the rubric in assessing students' SOAP note writing ability is indicated by significant positive correlations with preceptors' global impression of student performance and final APPE grades. Additionally, the use of this rubric allows for greater reliability among multiple graders and contributes to consistent grading and feedback throughout the APPE year. Although the time investment may be perceived as considerable by some preceptors, professional and concise SOAP note writing skills are essential for continuity of care in the primary care setting. Students should be provided opportunities to enhance these skills on APPEs. Using a standardized rubric can provide accurate and reliable SOAP note grades to strengthen student evaluations.

REFERENCES

1. Grainger-Rousseau TJ, Miralles MA, Hepler CD, Segal R, Doty RE, Ben-Joseph R. Therapeutic outcomes monitoring: application of pharmaceutical care guidelines to community pharmacy. *J Am Pharm Assoc.* 1997;NS37:647-661.
2. MacKinnon GE, III. Documentation: a value proposition for pharmacy education and the pharmacy profession. *Am J Pharm Educ.* 2007;71(4):Article 73.
3. Prosser TR, Burke JM, Hobson EH. Teaching pharmacy students to write in the medical record. *Am J Pharm Educ*. 1997;61:136-140.
4. Shogbon AO, Lundquist LM, Momary KM. Utilization of a structured approach to patient-based pharmacotherapy notes in a therapeutics course to improve clinical documentation skills. *Curr Pharm Teach Learn*. 2016;8(5):654-658.
5. Brown MC, Kotlyar M, Conway JM, Seifert R, St. Peter JV. Integration of an Internet-based medical chart into a pharmacotherapy lecture series. *Am J Pharm Educ.* 2007;71(3): Article 53.
6. Frenzel JE. Using electronic medical records to teach patient-centered care. *Am J Pharm Educ*. 2010;74(4):Article 71.
7. Sando KR, Skoy E, Bradley C, Fenzel J, Kirwin J, Urteaga E. Assessment of SOAP note evaluation tools in colleges and schools of pharmacy. *Curr Pharm Teach Learn.* 2017;9(4):576-584.
8. Conway JM, Ahmed GF. A pharmacotherapy capstone course to advance pharmacy students' clinical documentation skills. *Am J Pharm Educ.* 2012;76(7):Article 134.
9. Planas LG, Er NL. A systems approach to scaffold communication skills development. *Am J Pharm Educ.* 2008;72(2): Article 35.
10. Barnett SG, Gallimore C, Kopacek KJ, Porter AL. Evaluation of electronic SOAP note grading and feedback. *Curr Pharm Teach Learn.* 2014;6(4):516-526.
11. Accreditation Council for Pharmacy Education. Accreditation standards and key elements for the professional program in pharmacy leading to the doctor of pharmacy degree. Standards 2016. https://www.acpe-accredit.org/pdf/Standards2016FINAL.pdf. Accessed December 13, 2017.
12. Mertler CA. Designing scoring rubrics for your classroom. *Pract Assess Res Eval.* 2001;7(25).
13. Allen S, Knight J. A method for collaboratively developing and validating a rubric. *Int J Scholar Teach Learn*. 2009;3(2): Article 10.
14. Light RJ. Measures of response agreement for qualitative data: some generalizations and alternatives. *Psychol Bulletin*. 1971; 76(5):365-377.
15. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess*. 1994;6(4):284-290.
16. Agness CF, Huynh D, Brandt N. An introductory pharmacy practice experience based on a medication therapy management service model. *Am J Pharm Educ*. 2011;75(5):Article 82.
17. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Measure*. 1960;20(1):37-46.
18. Krippendorff K. *Content Analysis: An Introduction to Its Methodology*. Beverly Hills, CA: Sage Publications; 1980.

Downloaded from http://www.ajpe.org by guest on April 26, 2021. © 2018 American Association of Colleges of Pharmacy