Radiology

Predictors of Knowledge and Image Interpretation Skill Development in Radiology Residents[1]

Cécile J. Ravesloot, MD, PhD
Marieke F. van der Schaaf, PhD
Cas L. J. J. Kruitwagen, PhD
Anouk van der Gijp, MD
Dirk R. Rutgers, MD, PhD
Cees Haaring, BHS
Olle ten Cate, PhD
Jan P. J. van Schaik, MD, PhD

Purpose: To investigate knowledge and image interpretation skill development in residency by studying scores on knowledge and image questions on radiology tests, mediated by the training environment.

Materials and Methods: Ethical approval for the study was obtained from the ethical review board of the Netherlands Association for Medical Education. Longitudinal test data of 577 of 2884 radiology residents who took semiannual progress tests during 5 years were retrospectively analyzed by using a nonlinear mixed-effects model taking training length as input variable. Tests included nonimage and image questions that assessed knowledge and image interpretation skill. Hypothesized predictors were hospital type (academic or nonacademic), training hospital, enrollment age, sex, and test date.

Results: Scores showed a curvilinear growth during residency. Image scores increased faster during the first 3 years of residency and reached a higher maximum than knowledge scores (55.8% vs 45.1%). The slope of image score development versus knowledge question scores of 1st-year residents was 16.8% versus 12.4%, respectively. Training hospital environment appeared to be an important predictor in both knowledge and image interpretation skill development (maximum score difference between training hospitals was 23.2%; $P < .001$).

Conclusion: Expertise developed rapidly in the initial years of radiology residency and leveled off in the 3rd and 4th training year. The shape of the curve was mainly influenced by the specific training hospital.

Online supplemental material is available for this article.

[1] From the Department of Radiology (C.J.R., A.v.d.G., D.R.R., C.H., J.P.J.v.S.), Julius Center (C.L.J.J.K.) and Center for Research and Education Development (O.t.C.), University Medical Center Utrecht, Heidelberglaan 100, 3508 GA Utrecht, the Netherlands; and Department of Education, University Utrecht, Utrecht, the Netherlands (M.F.v.d.S.).
Received December 1, 2015; revision requested January 25, 2016 and received October 31; accepted November 22; final version accepted January 31, 2017.
Address correspondence to C.J.R. (e-mail: *C.J.Ravesloot@umcutrecht.nl*).

Radiologic knowledge and visual skills such as pattern recognition and efficient search strategies are fundamental for interpretation of radiologic images (1–3). Experienced radiologists possess a specialized knowledge and show advanced perceptual skills (4–7). However, increased experience does not automatically lead to higher image interpretation accuracy, and even experienced radiologists occasionally miss lesions (1,8,9). Improving image interpretation training in residency can reduce diagnostic errors. Insight in radiologic expertise development during residency and its predictors can help to shape the design of radiology training (1,10).

How expertise in radiology develops and which factors influence its development is largely unknown (10,11). In a previous cross-sectional study (12), we found that scores on knowledge and image questions in a radiology test differ between 1st-, 2nd-, and 3rd-year residents, but were similar for residents in postgraduate year 4 and 5. This suggested that knowledge and image interpretation skill development slows down during residency; however, this was not evaluated in a longitudinal study that tracked individual score development over time. Further, there are indications that specific hospital and resident factors might influence development of radiologic expertise (7,13–17).

Reading volume was a positive predictor of accuracy in mammogram interpretation (7,14) and might positively influence expertise development in radiology in general. Bhargavan and Sunshine (13) showed that the number of examinations per full-time equivalent radiologist in nonacademic hospitals was significantly larger than in academic hospitals in the United States. Consequently, image interpretation skill development of residents may be faster in nonacademic training programs than in academic programs.

Potential resident-related predictors of expertise development are sex and age. Women generally outperform men on knowledge tests and clinical examinations in undergraduate medical education (15,16). Kadmon et al (17) showed that higher enrollment age was associated with higher possibility of dropping out because of poor academic performance in German medical schools. Because of variation in previous clinical training or research activities, enrollment age differs among radiology residents. However, it is unknown whether this influences expertise development in radiology.

The aim of our study is to investigate knowledge and image interpretation skill development in residency by studying scores on knowledge and image questions on radiology tests, mediated by the training environment. Research questions are the following: *(a)* How does radiology expertise develop during residency? and *(b)* How does hospital type influence this development, controlled for sex and enrollment age?

We hypothesized that development of radiologic knowledge and image interpretation skill show a curvilinear growth (12). We also hypothesized that image interpretation skill develops faster in a nonacademic training environment because of higher reading volumes than in academic hospitals (13).

Advances in Knowledge

- Radiological knowledge rapidly increases during the initial phase of residency.
- Image interpretation skill develops faster than factual knowledge (16.8% vs 12.4% slope of score increase, respectively, during 1st training year).
- Radiological knowledge development in residency is influenced by training environment ($P < .001$).

Implication for Patient Care

- Predictors of expertise development may be used to improve radiology residency training; this might increase radiologic performance in clinical practice, and consequently improve patient care.

Materials and Methods

Study Design

Ethical approval for our study was obtained from the ethical review board of the Netherlands Association for Medical Education.

All data of 11 consecutive semiannual Dutch Radiology Progress Tests from 2005 to 2010 were retrospectively selected for our study with permission of the Radiological Society of the Netherlands. In November 2010, the test format changed, and therefore test results beyond November 2010 were excluded. If training length at the time of the test administration was inaccurate, this test result was excluded. For all other test results, percentage scores, training length, sex, age at enrollment, training hospital, and training hospital type (academic or nonacademic) were available. All data were anonymized and coded before they were transferred to the researchers. Scores on knowledge and image questions as a function of training length were analyzed to model the development of respectively radiologic knowledge and image interpretation skill, respectively. All possible predictors that were available were included in the analysis: hospital type, the coded individual training hospital, sex, enrollment age, and test date. The primary factors of interest were hospital type and training hospital. Sex, enrollment age, and test date were included in the model

https://doi.org/10.1148/radiol.2017152648

Content code: HP

Radiology 2017; 284:758–765

Author contributions:
Guarantors of integrity of entire study, C.J.R., M.F.v.d.S.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, C.J.R., A.v.d.G.; experimental studies, M.F.v.d.S., D.R.R., C.H., J.P.J.v.S.; statistical analysis, C.J.R., M.F.v.d.S., C.L.J.J.K., C.H.; and manuscript editing, C.J.R., M.F.v.d.S., C.L.J.J.K., A.v.d.G., D.R.R., J.P.J.v.S.

Conflicts of interest are listed at the end of this article.

to correct for potential influences. Test date was necessary to correct for difficulty differences between tests.

Sample and Population

Residents followed a 5-year training program; however, they did not all go straight through the program because of, for example, part-time work or research engagement. To control for these differences in pace, training length at the time of each test was calculated on a 5-year scale (for details on this calculation see Ravesloot et al [12]). For this calculation we used date at training enrollment, end of training date, date of the test, date of the previous test taken, and training length at time of the previous test. If training length could not be accurately determined, data were excluded from analysis (eg, when the overall training length shortened or the enrollment date changed during residency). If the total training duration for a resident appeared to be more than 7 years or less than 4.5 years, all test results pertaining to that resident were excluded. All test results of 31 residents (176 test results) and 20 test results of 11 residents were excluded. Residents frequently became exempt from taking a test once or more during their training for various reasons (eg, because of maternity leave or a doctorate trajectory) and therefore these results are missing. The median number of test results per resident in the dataset was 5.0 (interquartile range, 3.0–7.0). In total, 2884 test results from 11 tests from 577 residents in 26 hospitals were left for inclusion, 1477 of which pertained to residents in academic programs and 1407 to residents in nonacademic programs at the times of examination; 1779 scores (61.6%) were from men and 1105 (38.3%) were from women. There were 239 residents (41.4%) who stayed in nonacademic programs throughout their residency, 260 residents (45.1%) who stayed in academic programs, and 78 residents (13.5%) who switched between academic and nonacademic programs during their training. Median age at the start of residency training was 29.6 years (interquartile range, 27.8–31.4 years). The characteristics of training programs and their contribution to the dataset is shown in Table E1 (online).

Instrumentation

The Dutch Radiology Progress Test is a mandatory test simultaneously taken semiannually by all Dutch radiology residents and it covers all radiology subdomains. It is a progress test, which means that the same end-of-residency–level questions are answered by residents of all training levels. The Dutch Radiology Progress Test is a formative test that provides feedback for residents and program directors without a pass-or-fail decision. About 16% of the 200 questions included images (ie, image questions) that intend to test image interpretation skill. Images used in the test were conventional images and sections selected from computed tomographic (CT) scans and magnetic resonance (MR) images (ie, two-dimensional images). The remainder are nonimage questions that assess factual radiologic knowledge. According to Bloom taxonomy, factual knowledge involves “the basic elements that students must know to be acquainted with a discipline or solve problems in it” (18,19). All questions are true/false/do-not-know items. Scores are calculated by formula scoring (ie, subtracting incorrect from correct question scores) to correct for guessing (20). Answering “do not know” did not yield a score. Scores were calculated as a percentage from the maximum score. Questions with low reliability were removed from the test, and therefore maximum scores ranged from 188 to 197 (the total number of questions in the test). A detailed description of the test and its quality can be found in Ravesloot et al (12) and in Table E2 (online). In this previous study, test data from Dutch Radiology Progress Tests from 2005 to 2009 were evaluated for quality (12). Indicators of test quality were good. Reliabilities of the test scores were found to be consistently high (Cronbach α, ~0.90). Indications for construct validity of the test were found (eg, higher test scores with increased level of training).

Data Analyses

Model development.—Evaluation of raw data points of factual knowledge and image question scores (y-axis) versus training years (x-axis) demonstrated that the best-fitting curves showed a curvilinear shape for both factual knowledge and image question score development. Therefore, data were analyzed with a nonlinear mixed-effects model by using an asymptotic regression with an offset with three parameters, which approached the curvilinear shape at best as follows: *(a)* an asymptote that indicated the maximum score of an individual, *(b)* a starting point that indicated the number of training years at which an individual will have a score of 0, and *(c)* a log-rate constant that indicated the speed at which an individual’s score increases with training (Fig 1). These parameters are modeled by using a random intercept effect that allows for different values for each individual and accounts for dependence between repeated measurements on the individuals, and fixed predictor effects. Possible predictors were hospital, hospital type, sex, age at enrollment, and test date. Test date was included as a predictor to correct for difficulty differences between tests. If the contribution of hospital type (academic vs nonacademic) to the model was not statistically significant, hospital was included as possible predictor to test whether differences in scores were related to individual training hospitals. If hospital type was a significant predictor, hospital was added to the model to test for additional effects of differences related to individual training hospitals. No interactions between predictors were assessed because they lacked theoretical grounds.

Best-fitting models were obtained for knowledge and image scores separately. A stepwise modeling approach with the Akaike information criterion and convergence of the model was used to decide which variables to enter into or remove from the model. Akaike information criterion is on the basis of how well the model fits the data points corrected for the amount of degrees of freedom (21). By comparing

Figure 1

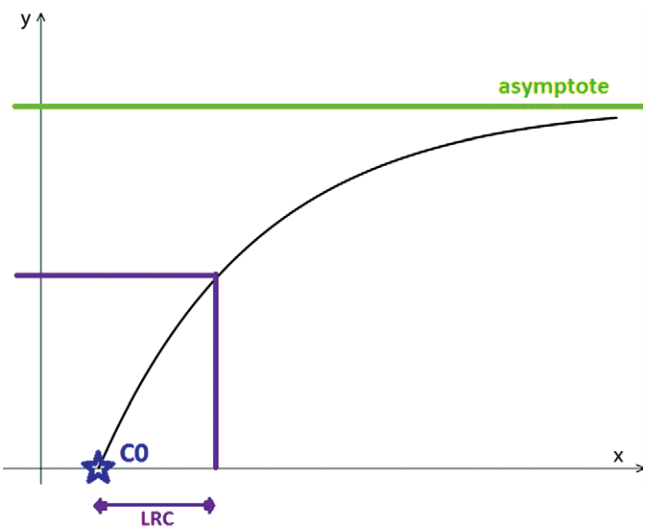


Figure 1: Graph of parameters in the applied nonlinear mixed-effects model. The x-axis is the training years and the y-axis is the test score. The green line is asymptote, which indicates the maximum score. The purple line crossing the y-axis shows when the participant reached half his maximum score (asymptote). The vertical purple line crossing the x-axis shows the number of training years the participant needed to reach 50% of his maximum score. The log-rate constant *(LRC)* indicates the speed at which half of the asymptote is reached. To measure this speed, the time needed to get from score zero to halfway the asymptotic score is used (purple line with double-ended arrows). Blue star is the time at score of 0 *(C0)*.

the two competing models, the model with the lowest Akaike information criterion is assumed to fit best. A variable was considered a statistically significant predictor when inclusion of this variable led to a decrease in the Akaike information criterion of the model. Subsequently, the influence of this predictor on each of the different separate parameters (asymptote, log-rate constant, and score of 0) was separately assessed again by using the Akaike information criterion. If inclusion of the variable led to an increase in Akaike information criterion, the variable was considered not significant and was removed from the model. Data were analyzed with the SSAsympOff function of the library “nlme” of a statistical program (R Version 2.15.2; R Foundation for Statistical Computing, Vienna, Austria).

Test results from residents who were trained in both academic and nonacademic programs did not influence the model or the results of the study, and were included in the analyses.

Modeling the predictors of expertise development.—To estimate the speed of knowledge and image interpretation skill development during residency, the slope of the predicted curves of the knowledge and image score best-fitted models were calculated at four time points: 1, 2, 3, and 4 years after start of residency.

Coefficients of the statistically significant predictors in both models were used to estimate the magnitude of the influence of each predictor on expertise regarding knowledge or image interpretation skill development.

Results

Predictors of Factual Radiologic Knowledge Development

The best-predicting model of factual knowledge development of the average subject is shown in Figure 2a and represents knowledge question score development during residency. Model characteristics are shown in the Table. The model included test date, hospital, and age at enrollment, and explained 86.4% of the variance in factual knowledge score within individuals. The predicted average factual knowledge score was 10.9% at the start of residency and 45.1% at the end of residency. The slope of the curve decreased 12.4%, 6.3%, 3.2%, and 1.6% per year after 1, 2, 3, and 4 years of training, respectively (ie, 50% every year; Fig 2a).

The maximum score was influenced by the specific training hospital, which also significantly influenced the slope of the curve. The difference in average predicted maximum score between individuals in lowest- and highest-scoring hospitals was 22.9% (36.5%–59.4%; Fig 3a). Age at enrollment was also a significant predictor of the asymptote. The average predicted maximum percentage score increased 0.72 (standard error, 0.20; $P < .001$) per year with participants who were younger at enrollment (Fig 3b). Hospital type (academic or nonacademic) and sex did not significantly improve the model of factual knowledge development and were therefore not included in the model.

Predictors of Radiologic Image Interpretation Skill Development

The model that best predicted image interpretation skill development during residency included specific training hospital, age at enrollment, and sex (Fig 2, Table). The model explained 67.9% of the variance in image score development within individuals. Image interpretation development had the same curvilinear pattern as factual knowledge question score development but increased faster and reached a higher predicted score at end of residency (55.8%). The slope of the curve after 1, 2, 3, and 4 years of training was 16.8%, 8.2%, 4.0%, and 2.0% per year, respectively (ie, it decreased by 50% every year).

The maximum image question score was significantly influenced by

Radiology

Figure 2

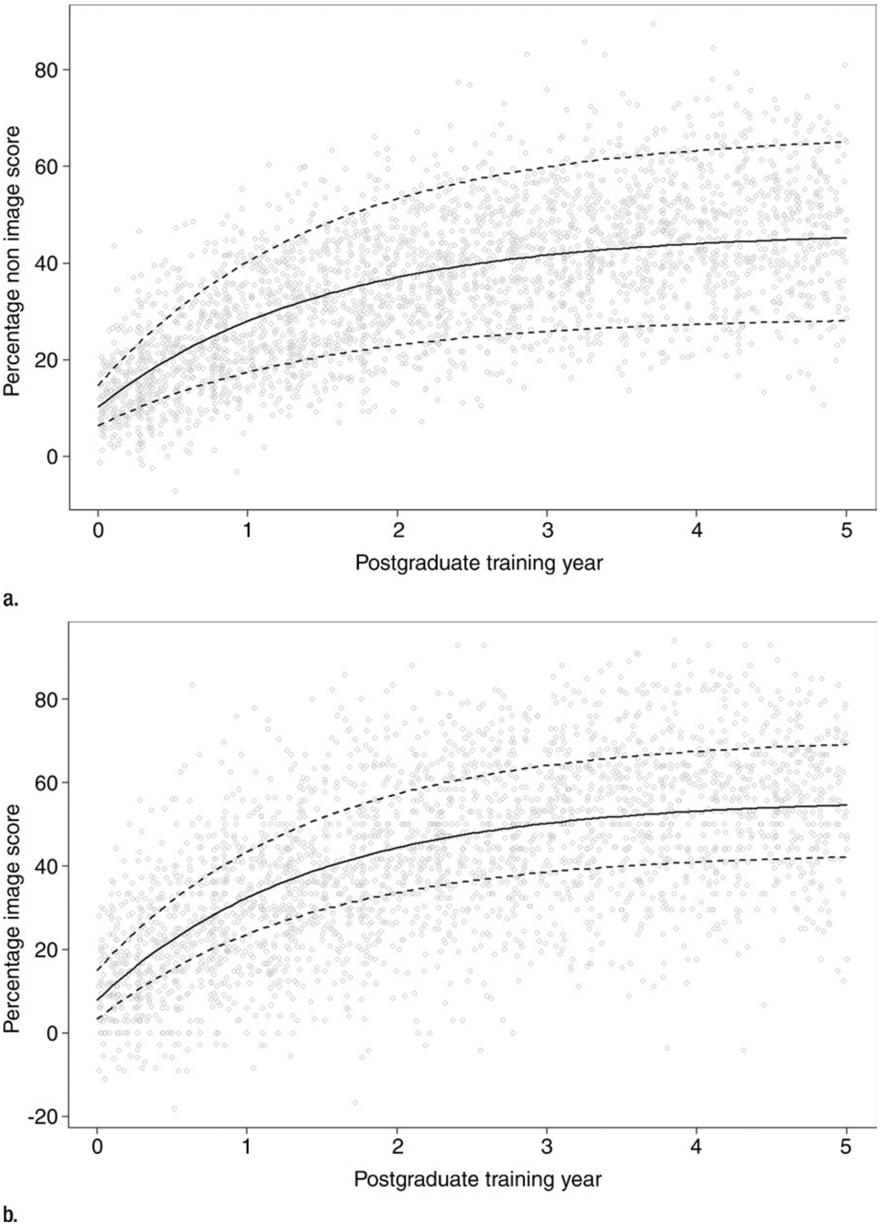


Figure 2: Graphs show predicted curve for a fictitious individual with mean values for covariates against training length, plus curves (dotted lines) indicating where the middle 95% of individuals would have scored if they had mean covariate values. Longitudinal development of **(a)** factual knowledge and **(b)** image interpretation skill scores during radiology residency on the basis of the following equations for knowledge-based (*nonimage*) and image-based question scores, respectively: $y = 46.4 \times \{1 - \exp[-0.68 \times (x + 0.37)]\}$ and $y = 55.8 \times \{1 - \exp[-0.71 \times (x + 0.21)]\}$, where *exp* is exponent. All parameter estimates are significant, with *P* values less than .001 (Table).

Characteristics of Best-fitting Models for Knowledge and Image Question Scores

Predictor	*P* Value
Knowledge score model	
Asymptote	
Test date	<.001
Age at enrollment	<.001
Training hospital	<.001
Log rate constant	
Training hospital	< .005
Starting point C0	
Test date	<.005
Image score model	
Asymptote	
Test date	<.001
Age at enrollment	<.001
Training hospital	<.001
Log rate constant	
Sex	<.01

Note.—Scores are based on the Akaike information criterion. C0 = score of 0. *P* values obtained from an analysis-of-variance table.

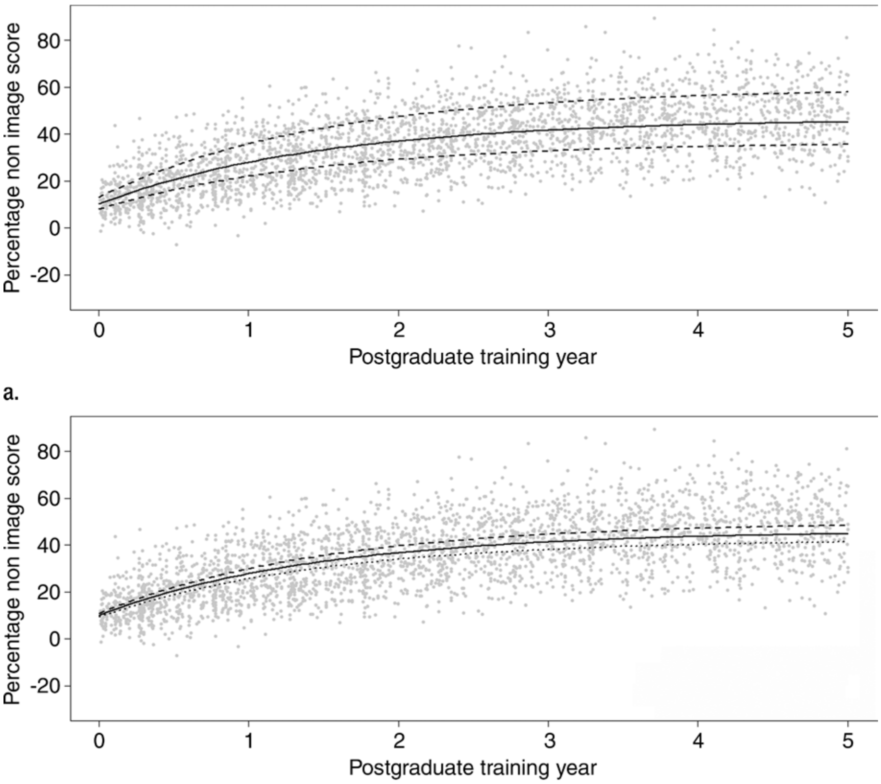
the specific training hospital and age at enrollment. The average predicted score of individuals in lowest-scoring and highest-scoring hospitals was 45.4% and 68.6%, respectively (Fig 4). For every year younger at time of enrollment, the average predicted percentage image interpretation skill score at end of residency was 1.0 higher (standard error, 0.22; $P < .001$).

Sex influenced the slope of image question score development but did not influence the asymptote of image question score.

Discussion

Factual radiologic knowledge and image interpretation skill development as reflected by knowledge and image question score development showed a curvilinear shape during residency. After 1 year of training, the slope of the knowledge and image question scores measured 12.4% and 16.8%, respectively, and every following year the slope decreased by 50% until it reached 1.6% and 2.0%, respectively, after 4 years of training. This aligns with the generic Thurstone learning curve and with other empirical examples that show that learning gain decelerates over time (22–25). In particular, scores on progress testing in undergraduate medical education

Figure 3



b.

Figure 3: Predicted curves of development of scores on factual knowledge questions for individuals **(a)** in training hospitals with average (continuous line), lowest (dotted line), and highest (dotted line) scores and **(b)** with enrollment at age 25 (dashed line), 30 (solid line), and 35 (dotted line) years and otherwise characteristics for an average fictitious individual in the dataset.

Figure 4

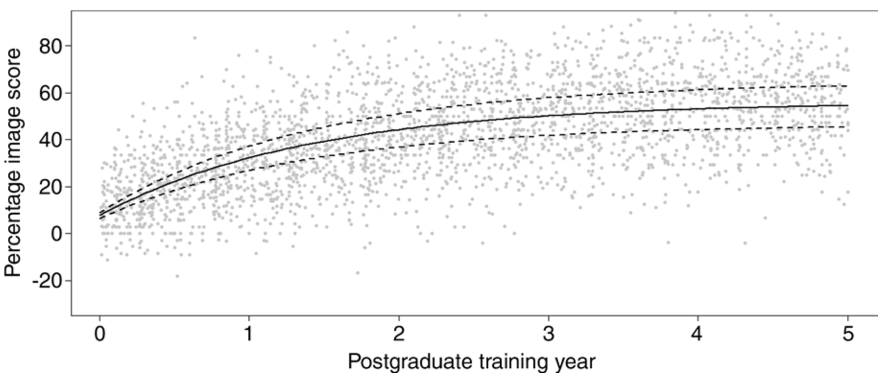


Figure 4: Predicted curves of development of scores on image questions for residents in training hospitals with average (continuous line), lowest (dotted line), and highest (dotted line) scores, and otherwise average characteristics.

showed similar decelerating growth of knowledge (20,26). Our findings also align with results on a study regarding skeletal radiograph interpretation development that shows a short phase of rapid increase of resident performance, after which learning slows down (27).

This result has two consequences for radiologic training. First, individual progress evaluation in the first 3 years of residency is important. Because much progression is expected in this period, below-average improvement in an early phase can be an indication for concern and may require in-depth evaluation of clinical performance and additional training. Second, in the final training year, experience alone might not be sufficient to further increase expertise. For example, more focused training and feedback on knowledge gaps or image interpretation errors might be important after the concept of deliberate practice, which implies that expertise in a specific domain develops best as a result of repetitive execution of relevant tasks with specific and immediate feedback and reflection (28).

From the beginning of residency, the acquisition of radiologic image interpretation skill appeared to develop at a faster rate than factual knowledge scores. This aligned with previous findings (14) and implied that radiology expertise begins to develop directly from the start of exposure to images. It suggested that not much radiology-specific factual knowledge was needed for this initial development.

We hypothesized that, because of a high workload and image exposure in nonacademic hospital contexts, this setting would yield a more rapid image interpretation skill development. Our findings do not confirm this. One possible explanation might be that a lower workload in academic hospitals allows for more in-depth study and receipt of feedback, and thus for deliberate practice and compensation for the lower image exposure in this setting (13,29). Further, there may be a limit to the effective amount of image exposure for learning image interpretation skill. Another explanation is that the Dutch

Radiology Progress Test includes questions on less prevalent illnesses, which are rare in nonacademic hospitals. Unlike hospital type, hospital was a statistically significant predictor of knowledge and image interpretation skill development. Application for residency in the Netherlands happens through an open-market competitive selection procedure. Criteria are defined by individual training hospitals and differ from site to site. Because the data were anonymous for the hospital, it is unclear what caused these differences; they could be patient populations, resident selection procedures, educational programs, or other factors.

On the basis of our model, trainees who begin radiology residency directly after medical school are expected to show the largest development in knowledge and image interpretation skill, which results in slightly higher final scores. This is consistent with the sparse evidence on this topic in undergraduate medical education (17). An explanatory note regarding transition from university medical education to graduate medical education in the Netherlands is needed. The majority of medical graduates do not start residency immediately after medical school, but spend time to acquire additional clinical experience, prepare for applications, perform research, or spend months to years otherwise before starting postgraduate training.

Other than the slightly faster initial increase of image scores for male participants, sex differences did not influence expertise development. This contradicts results from studies that show that women generally outperform men in medical school (15,16). One explanation might be that, on average, men have better spatial abilities than women (30), which might be positively related to initial radiologic image interpretation development (31,32).

Our results can help to improve the training of radiology residents in several ways. First, it is important to realize that factual radiologic knowledge and image interpretation skill development are related to training environment. Research that explores differences in training hospitals and their individual influence on development of expertise may be useful for designing successful training environments. Second, at the end of training, experience alone seems not sufficient for expertise growth. More research should be focused on successful methods for continuous growth. Third, evaluation in the initial years of residency is important because knowledge growth is expected to be largest during that period.

Our study has limitations. First, the Dutch Radiology Progress Test does not measure the full range of requisite skills in clinical practice. For one thing, image questions were on the basis of only two-dimensional images (ie, conventional images and single sections selected from CT scans and MR images) (33–35). Stack viewing and other advanced image manipulation skills were not tested, and neither were practical skills (eg, ultrasonography and interventions) nor communication skills. Second, residents were not randomly distributed over hospitals because hospitals use different selection procedures. Effects of hospital on expertise development might therefore be biased by resident differences. Third, we did not measure the effect of workload on expertise development directly, but indirectly inferred this by assuming that workload, in the sense of patient caseload, was higher in nonacademic hospitals than in academic hospitals.

Expertise developed rapidly in the initial years of radiology residency and leveled off in the 3rd and 4th training year. The shape of the curve was mainly influenced by the specific training hospital.

Acknowledgments: The authors thank the members of the examination committee of the Dutch Radiology Progress Test and the Radiological Society of the Netherlands for their support.

Disclosures of Conflicts of Interest: **C.J.R.** disclosed no relevant relationships. **M.F.v.d.S.** disclosed no relevant relationships. **C.L.J.J.K.** disclosed no relevant relationships. **A.v.d.G.** disclosed no relevant relationships. **D.R.R.** disclosed no relevant relationships. **C.H.** disclosed no relevant relationships. **O.t.C.** disclosed no relevant relationships. **J.P.J.v.S.** disclosed no relevant relationships.

References

1. Taylor PM. A review of research into the development of radiologic expertise: implications for computer-based training. Acad Radiol 2007;14(10):1252–1263.
2. van der Gijp A, van der Schaaf MF, van der Schaaf IC, et al. Interpretation of radiological images: towards a framework of knowledge and skills. Adv Health Sci Educ Theory Pract 2014;19(4):565–580.
3. Drew T, Evans K, Võ ML, Jacobson FL, Wolfe JM. Informatics in radiology: what can you see in a single glance and how might this guide visual search in medical images? RadioGraphics 2013;33(1):263–274.
4. Gegenfurtner A, Lehtinen E, Saljo R. Expertise differences in the comprehension of visualizations: a meta-anaysis of eye-tracking research in professional domains. Educ Psychol Rev 2011;23(4):523–552.
5. Kundel HL, Nodine CF, Conant EF, Weinstein SP. Holistic component of image perception in mammogram interpretation: gaze-tracking study. Radiology 2007;242(2):396–402.
6. Mallett S, Phillips P, Fanshawe TR, et al. Tracking eye gaze during interpretation of endoluminal three-dimensional CT colonography: visual perception of experienced and inexperienced readers. Radiology 2014;273(3):783–792.
7. Nodine CF, Kundel HL, Mello-Thoms C, et al. How experience and training influence mammography expertise. Acad Radiol 1999;6(10):575–585.
8. Berlin L. Radiologic errors, past, present and future. Diagnosis 2014;1(1):79–84.
9. Krupinski EA. Current perspectives in medical image perception. Atten Percept Psychophys 2010;72(5):1205–1217.
10. Gunderman R, Williamson K, Fraley R, Steele J. Expertise: implications for radiological education. Acad Radiol 2001;8(12):1252–1256.
11. Krupinski EA. The role of perception in imaging: past and future. Semin Nucl Med 2011;41(6):392–400.
12. Ravesloot C, van der Schaaf M, Haaring C, et al. Construct validation of progress testing to measure knowledge and visual skills in radiology. Med Teach 2012;34(12):1047–1055.
13. Bhargavan M, Sunshine JH. Workload of radiologists in the United States in 2002-2003 and trends since 1991-1992. Radiology 2005;236(3):920–931.
14. Esserman L, Cowley H, Eberle C, et al. Improving the accuracy of mammography:

volume and outcome relationships. J Natl Cancer Inst 2002;94(5):369–375.

15. Cuddy MM, Swanson DB, Dillon GF, Holtman MC, Clauser BE. A multilevel analysis of the relationships between selected examinee characteristics and United States Medical Licensing Examination Step 2 Clinical Knowledge performance: revisiting old findings and asking new questions. Acad Med 2006;81(10 Suppl):S103–S107.

16. Cuddy MM, Swanson DB, Clauser BE. A multilevel analysis of examinee gender and USMLE step 1 performance. Acad Med 2008;83(10 Suppl):S58–S62.

17. Kadmon G, Resch F, Duelli R, Kadmon M. Predictive value of the school-leaving grade and prognosis of different admission groups for academic performance and continuity in the medical course - a longitudinal study. GMS Z Med Ausbild 2014;31(2):Doc21.

18. Bloom BS, Engelhart MD, Furst EJ, Hill WH, Krathwohl DR. Taxonomy of educational objectives: the classification of educational goals. New York, NY: McKay, 1956.

19. Anderson LW, Krathwohl DR, Airasian PW, et al. A taxonomy for learning, teaching, and assessing: a revision of Bloom's taxonomy of educational objectives. New York, NY: Addison Wesley Longman, 2001.

20. McHarg J, Bradley P, Chamberlain S, Ricketts C, Searle J, McLachlan JC. Assessment of progress tests. Med Educ 2005;39(2):221–227.

21. Akaike H. A new look at the statistical model identification. IEEE Trans Automat Contr 1974;19(6):716–723.

22. Thurstone LL. The learning curve equation. Princeton, NJ: Psychological Review Company, 1919.

23. ten Cate O, Snell L, Carraccio C. Medical competence: the interplay between individual ability and the health care environment. Med Teach 2010;32(8):669–675.

24. Konrad C, Schüpfer G, Wietlisbach M, Gerber H. Learning manual skills in anesthesiology: Is there a recommended number of cases for anesthetic procedures? Anesth Analg 1998;86(3):635–639.

25. Beck MR, Martin BA, Smitherman E, Gaschen L. Eyes-on training and radiological expertise: an examination of expertise development and its effects on visual working memory. Hum Factors 2013;55(4):747–763.

26. Blake JM, Norman GR, Keane DR, Mueller CB, Cunnington J, Didyk N. Introducing progress testing in McMaster University's problem-based medical curriculum: psychometric properties and effect on learning. Acad Med 1996;71(9):1002–1007.

27. Pusic M, Pecaric M, Boutis K. How much practice is enough? Using learning curves to assess the deliberate practice of radiograph interpretation. Acad Med 2011;86(6):731–736.

28. Ericsson KA. Deliberate practice and acquisition of expert performance: a general overview. Acad Emerg Med 2008;15(11):988–994.

29. Ericsson KA. An expert-performance perspective of research on medical expertise: the study of clinical performance. Med Educ 2007;41(12):1124–1130.

30. Voyer D, Voyer S, Bryden MP. Magnitude of sex differences in spatial abilities: a meta-analysis and consideration of critical variables. Psychol Bull 1995;117(2):250–270.

31. Berbaum KS, Smoker WR, Smith WL. Measurement and prediction of diagnostic performance during radiology training. AJR Am J Roentgenol 1985;145(6):1305–1311.

32. Nilsson T, Hedman L, Ahlqvist J. Visual-spatial ability and interpretation of three-dimensional information in radiographs. Dentomaxillofac Radiol 2007;36(2):86–91.

33. Ravesloot CJ, van der Schaaf MF, van Schaik JP, et al. Volumetric CT-images improve testing of radiological image interpretation skills. Eur J Radiol 2015;84(5):856–861.

34. Ravesloot CJ, van der Gijp A, van der Schaaf MF, et al. Support for external validity of radiological anatomy tests using volumetric images. Acad Radiol 2015;22(5):640–645.

35. van der Gijp A, Ravesloot CJ, van der Schaaf MF, et al. Volumetric and two-dimensional image interpretation show different cognitive processes in learners. Acad Radiol 2015;22(5):632–639.