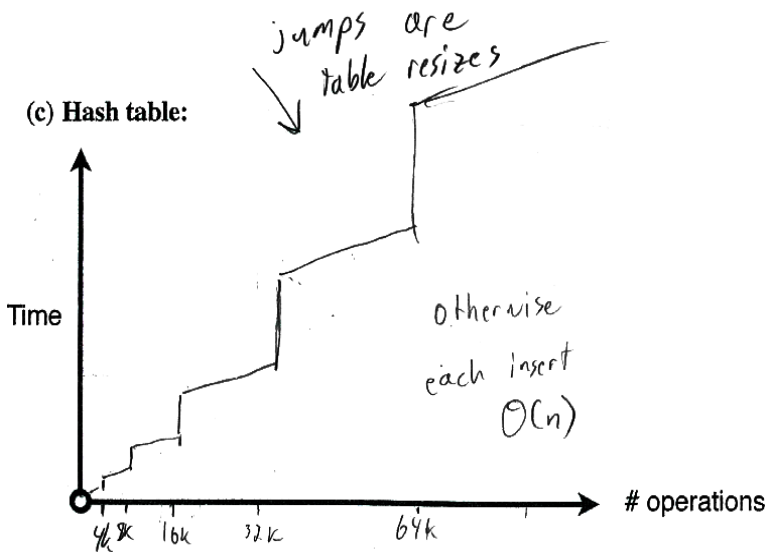
Misconceptions and Concept Inventory Questions for Binary Search Trees and Hash Tables

Kuba Karpierz, Steven Wolfman,
and others in the CS-SEI

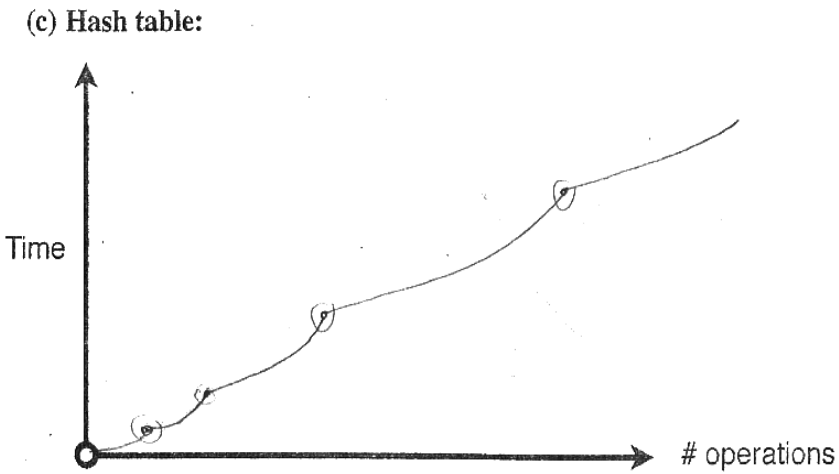
Concept Inventory Status

Identifying key learning goals	Instructor interviews	
Studying student artifacts	Analysis of 200+ Qs (15+ exams) + project submissions	**Still working on a few key topics**
Designing open-ended questions	~30 draft Qs	
Think-aloud interviews	~25 hrs of interviews	
Designing MC questions	~25 draft Qs	
Validation interviews	~5 hrs of interviews	***Much* more needed (tricky with iterative development!)**
Expert validation	Feedback @ broad presentation + some 1-1 feedback	
Data collection and analysis	9 offerings (~600 students); much analysis left to go!	**Much analysis & feedback left to do!**
Feedback to instructors & curriculum	Dep't talk & 1-1 with instructors in area	

Correct Answer



Common Misconception



CPSC 221 Exam Answers

Which of these **best** describes the process of resizing a hash table to increase its size?

a. Create a new, separate table larger than the old table, and copy the keys from the old table to corresponding slots in the new one.

b. Create a new, separate table larger than the old table, and individually re-insert each key from the old table into the new one.

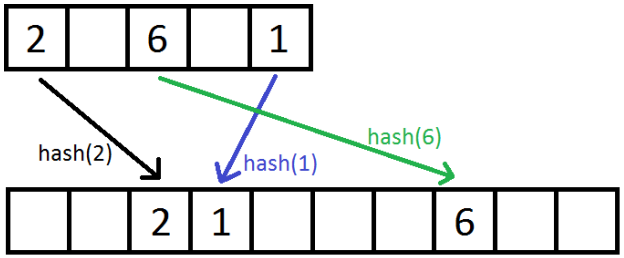
c. Add extra slots at the end of the existing hash table, and leave the old keys in place in their existing slots.

d. Add extra slots at the end of the existing hash table, and individually re-insert each key into the table.

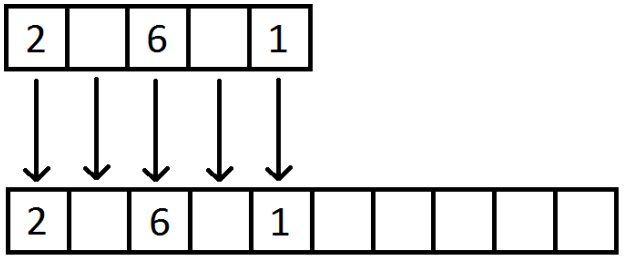
Concept Inventory Question

Artifact analysis, student think-alouds, expert focus group, pilot testing and validation think-alouds

Rehashing vs. Block-Copying

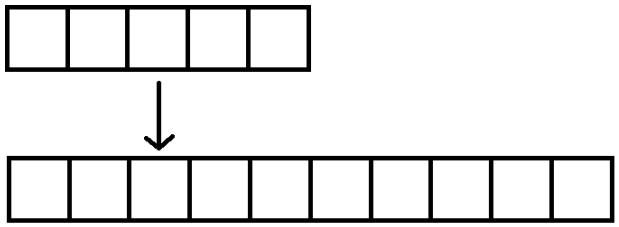


vs.

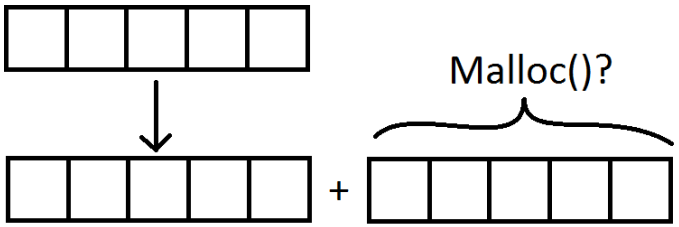


CPSC 221	CPSC 320
38%	23%

Reallocating vs. Extending



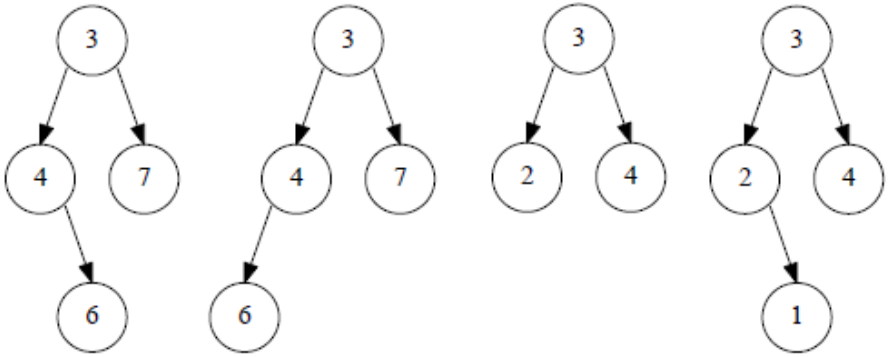
vs.



CPSC 221	CPSC 320
22%	23%

(Percentages of students who answered each question.)

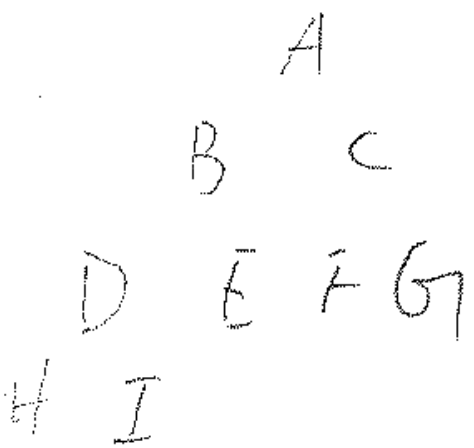
For each of these: is it
a heap (only), BST (only),
both, or neither?

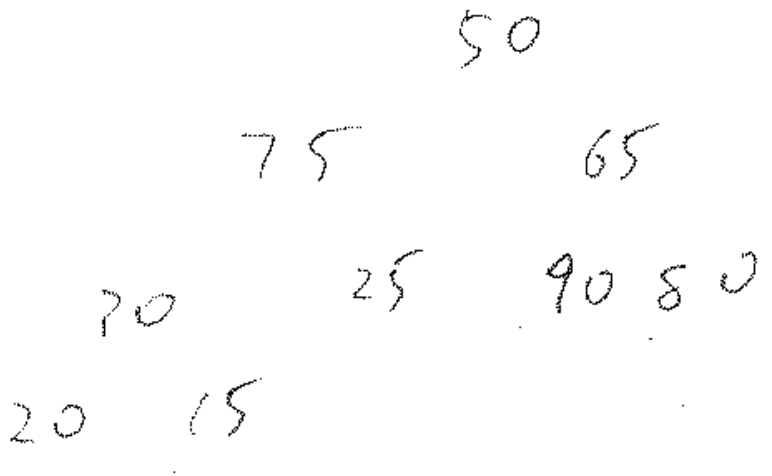


BST/Heap conflation in prior work (Danielsiek et al., SIGCSE 2012)
Unable to replicate (us or original authors)

And yet, similar misconception appeared in think-alouds, exam analysis, etc.

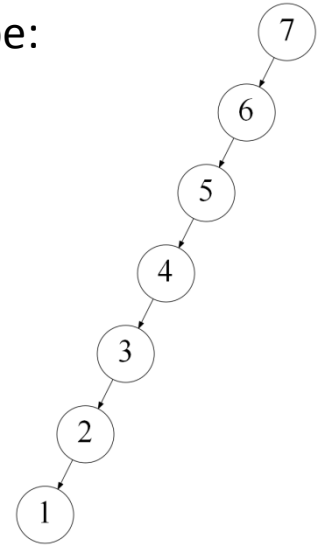
Draw a BST whose keys printed in post-order traversal are: 20 15 30 25 75 90 80 65 50.



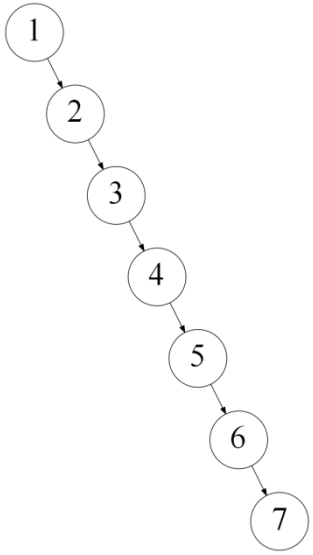


What shape is a binary search tree that contains the keys 1, 2, 3, 4, 5, 6, and 7?

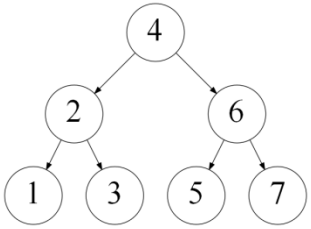
a. Exactly this shape:



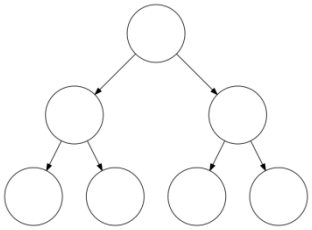
c. Exactly this shape:



b. Exactly this shape:



d. This shape with either 1 or 7 at the root:



e. There is not enough information to tell.

Think-alouds: "I just want to pick (b) because it looks the nicest."
"I'm going to assume they're looking for the perfect BST that contains said keys."

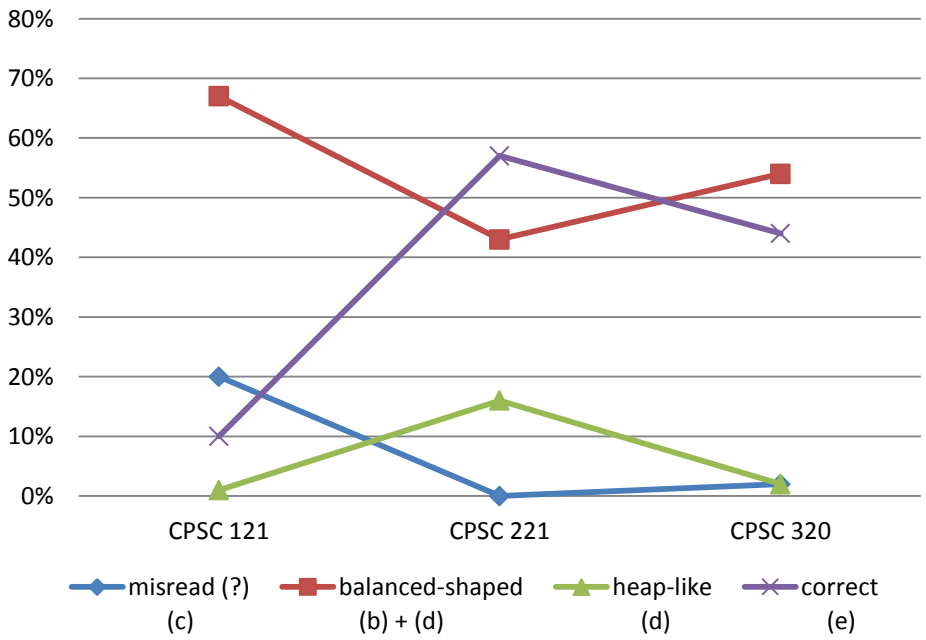
Expert feedback → "A binary search tree contains the keys 1, 2, 3, 4, 5, 6, and 7.
What shape **must** the tree be?"

CPSC 121 (and CPSC 110 co-req): BSTs illustrate an interesting recursive structure. Little discussion of efficiency or visualization of algorithms.

CPSC 221: BSTs discussed extensively, implemented, and used as foundation for a variety of other data structures. Binary trees (**not** BSTs) used as foundation of heap data structure during same term.

CPSC 320: Continued use of trees as analysis tool and data structure. (Little continued use and less continued study of heap data structure.)

	CPSC 121	CPSC 221	CPSC 320
a	3%	0%	0%
b	66%	27%	52%
c	20%	0%	2%
d	1%	16%	2%
e	10%	57%	44%



Other Current CI Questions

- Data collection/analysis stage
  - Determining what a proof means
  - Classifying functions as exponential
  - Describing code with recurrence relations
  - …
- Earlier stages
  - Induction
  - Dynamic Programming
  - …